



Misinformation, Platforms, and the Role of Reasoning

*How Do We Control Misinformation?
It Depends on Reasoning Abilities*

MOHAMED MOSTAGIR

JAMES SIDERIUS

*Misinformation: Strategic Sharing, Homophily,
and Endogenous Echo Chambers*

DARON ACEMOGLU

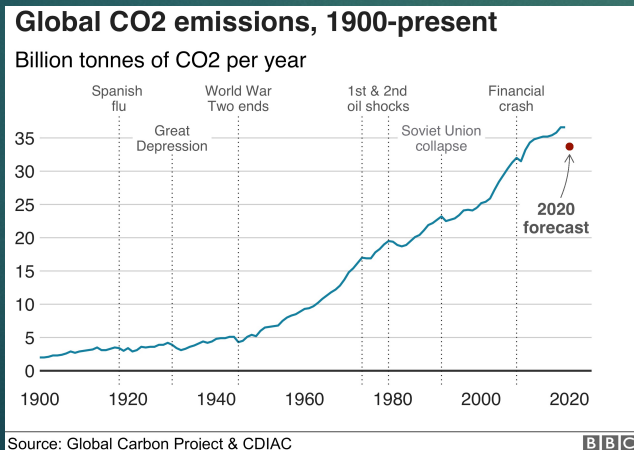
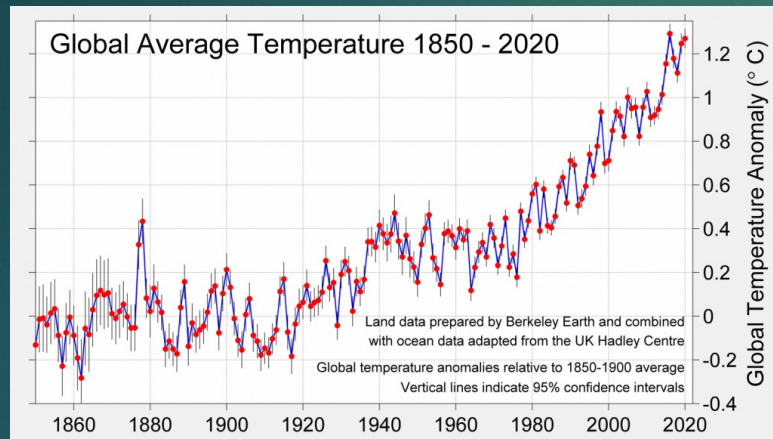
ASU OZDAGLAR

JAMES SIDERIUS

Organic News vs. Misinformation

Organic

Misinformation



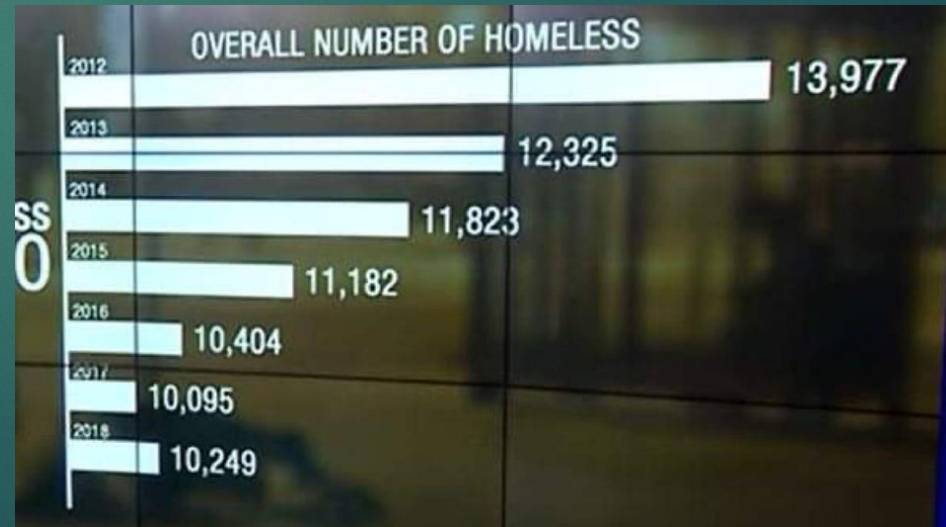
WESTERNJOURNAL.COM

Dem's Vote To Enhance Med Care for Illegals Now, Vote Down Vets Waiting 10 Years for Same Service

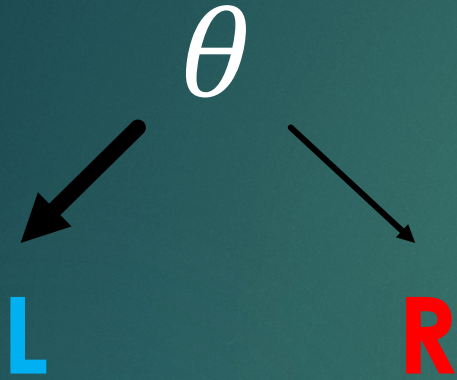
POLITICUSUSA.COM | BY JASON EASLEY

Trump Is Now Trying To Get Mike Pence Impeached

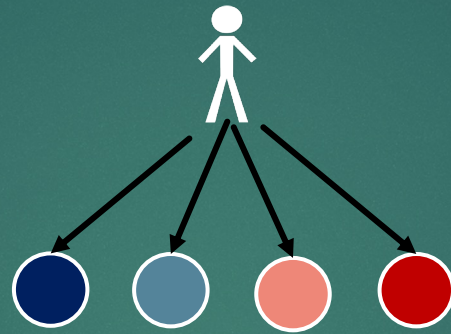
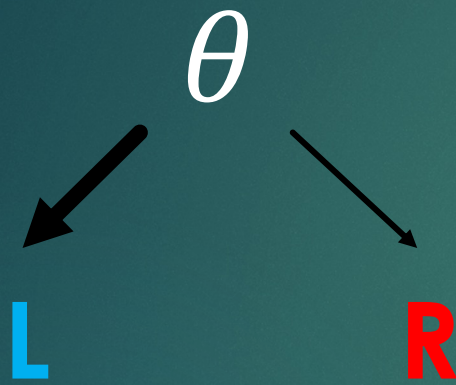
During a press conference, Trump said that if he is going to be...



Model: Network of Social Interactions

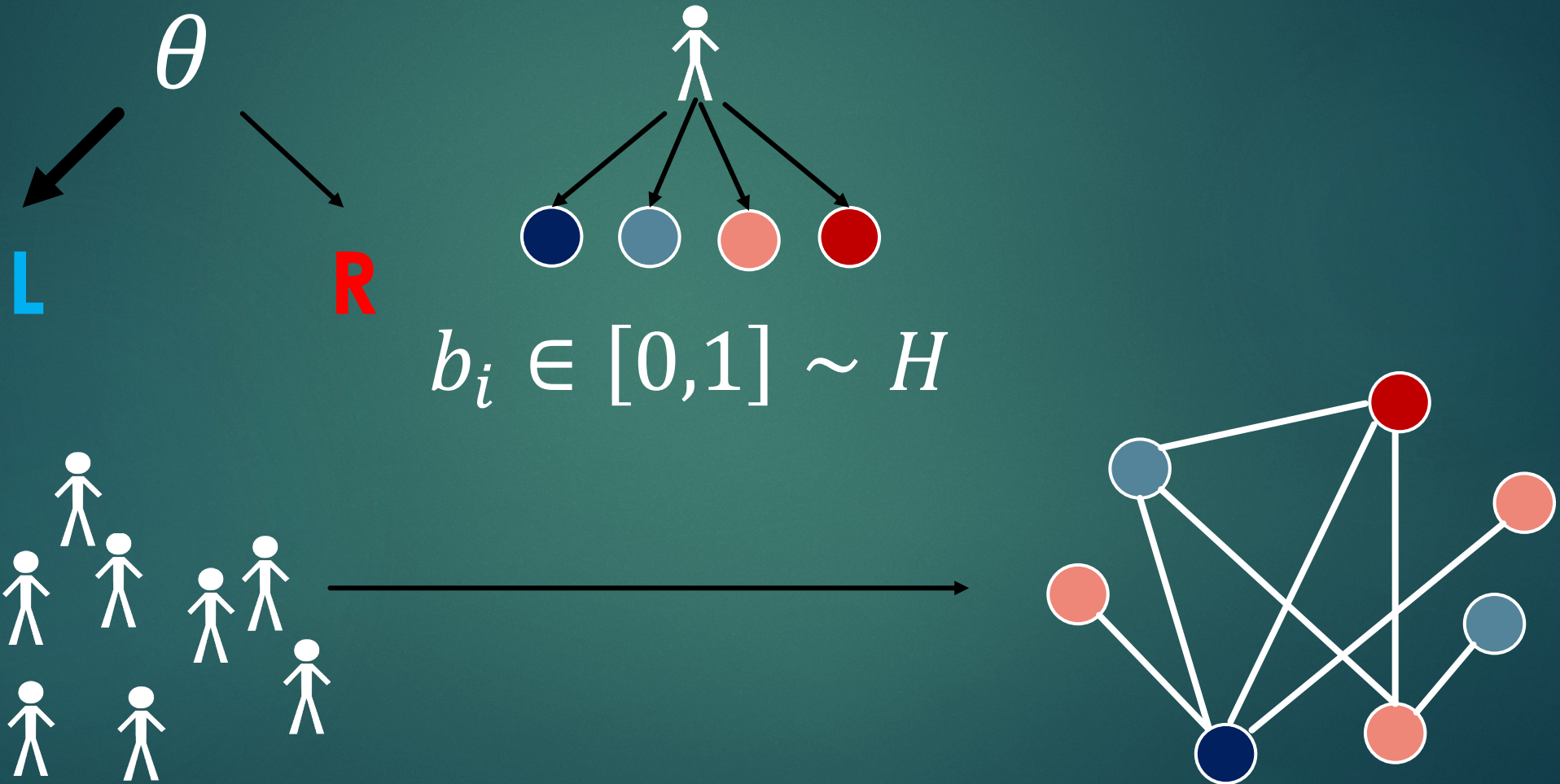


Model: Network of Social Interactions



$$b_i \in [0,1] \sim H$$

Model: Network of Social Interactions



Two Models: Learning and Sharing

- ▶ Misinformation is bad for communicating the ground truth.

Two Models: Learning and Sharing

- ▶ Misinformation is bad for communicating the ground truth.
- ▶ **Mostagir and Siderius (2021):** *How Do We Control Misinformation? It Depends on Reasoning Abilities*
 - ▶ Share beliefs: agents consume content but then share their opinions over social media.
 - ▶ Two sophistication types: **sophisticated** and **unsophisticated**. Update beliefs based on content and learn from other beliefs differently.

Two Models: Learning and Sharing

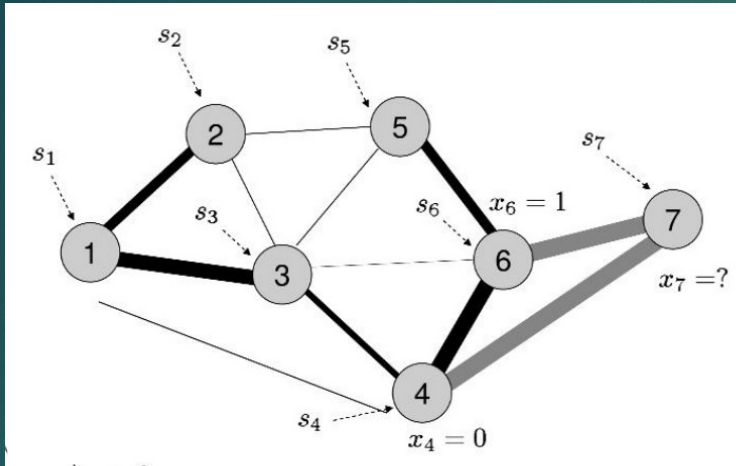
- ▶ Misinformation is bad for communicating the ground truth.
- ▶ **Mostagir and Siderius (2021):** *How Do We Control Misinformation? It Depends on Reasoning Abilities*
 - ▶ Share beliefs: agents consume content but then share their opinions over social media.
 - ▶ Two sophistication types: **sophisticated** and **unsophisticated**. Update beliefs based on content and learn from other beliefs differently.
- ▶ **Acemoglu, Ozdaglar, and Siderius (2021):** *Misinformation: Strategic Sharing, Homophily, and Endogenous Echo Chambers*
 - ▶ Share content: agents choose whether to pass content onto others.
 - ▶ When does misinformation spread?



How Do We Control Misinformation? It Depends on Reasoning Abilities

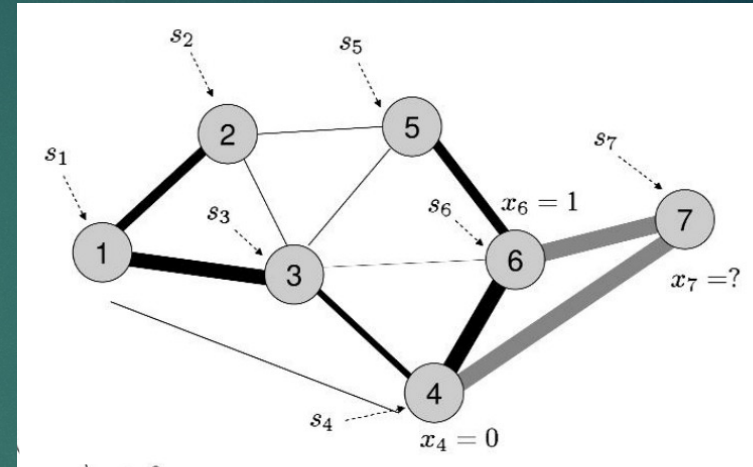
MODEL OF LEARNING

Classical Learning



[Sophisticated]

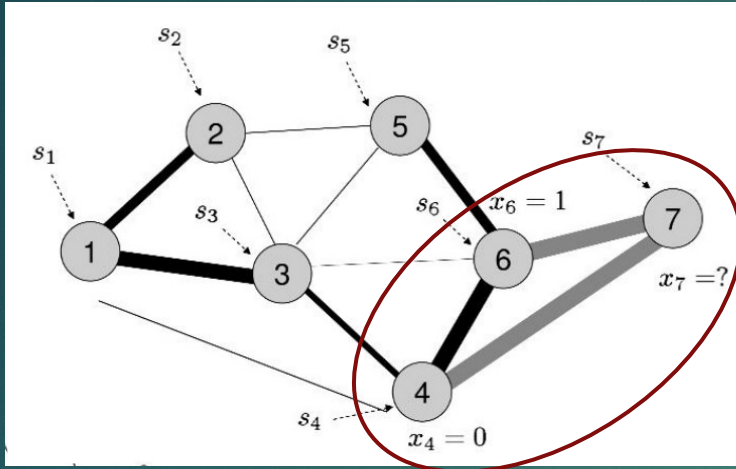
e.g., Acemoglu et al (2011): Bayesian Learning in Social Networks



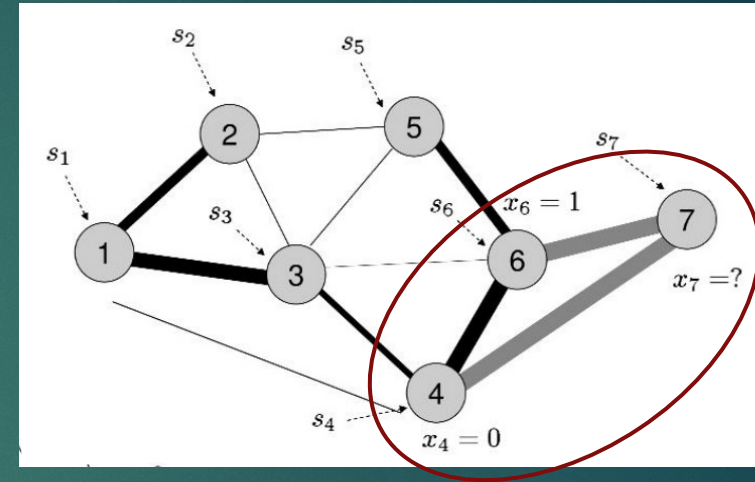
[Unsophisticated]

e.g., Golub et al (2010): Naïve Learning in Social Networks and the Wisdom of Crowds

Classical Learning

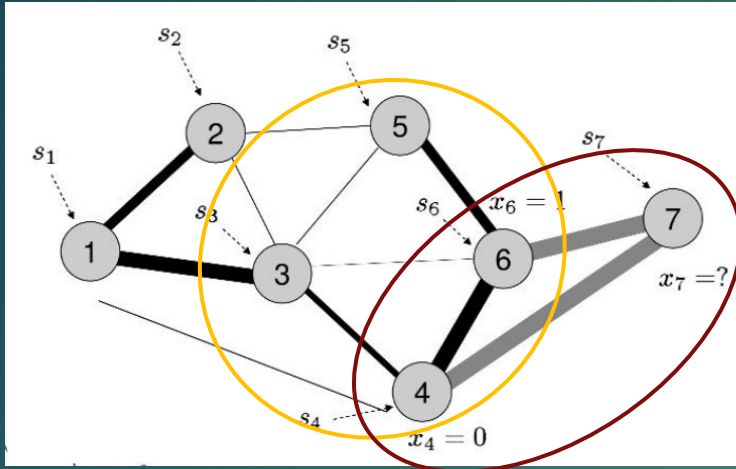


[Sophisticated]
e.g., Acemoglu et al (2011): Bayesian Learning in Social Networks

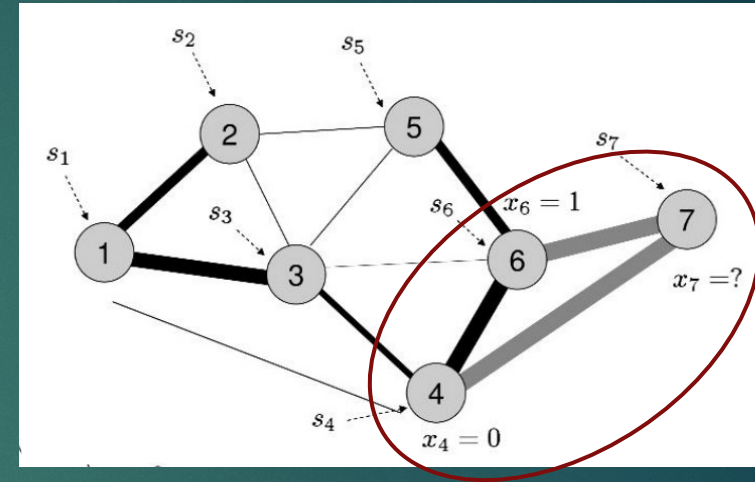


[Unsophisticated]
e.g., Golub et al (2010): Naïve Learning in Social Networks and the Wisdom of Crowds

Classical Learning

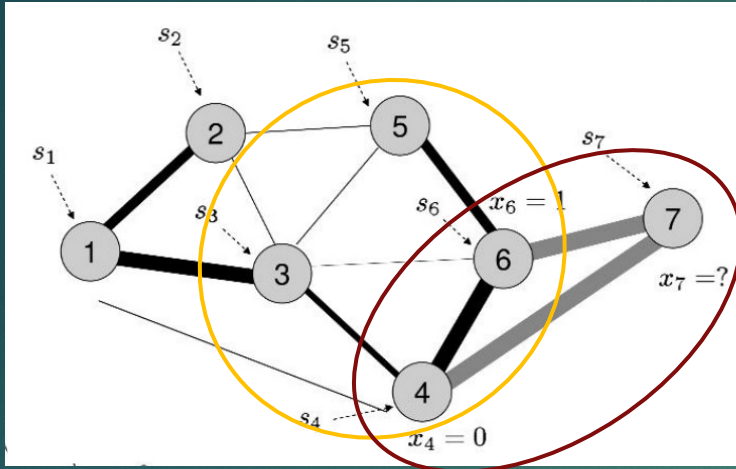


[Sophisticated]
e.g., Acemoglu et al (2011): Bayesian
Learning in Social Networks



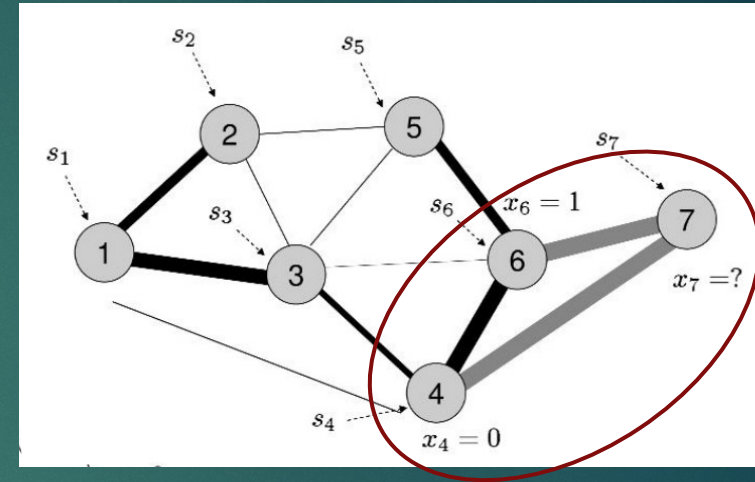
[Unsophisticated]
e.g., Golub et al (2010): Naïve Learning in
Social Networks and the Wisdom of Crowds

Classical Learning



[Sophisticated]
e.g., Acemoglu et al (2011): Bayesian
Learning in Social Networks

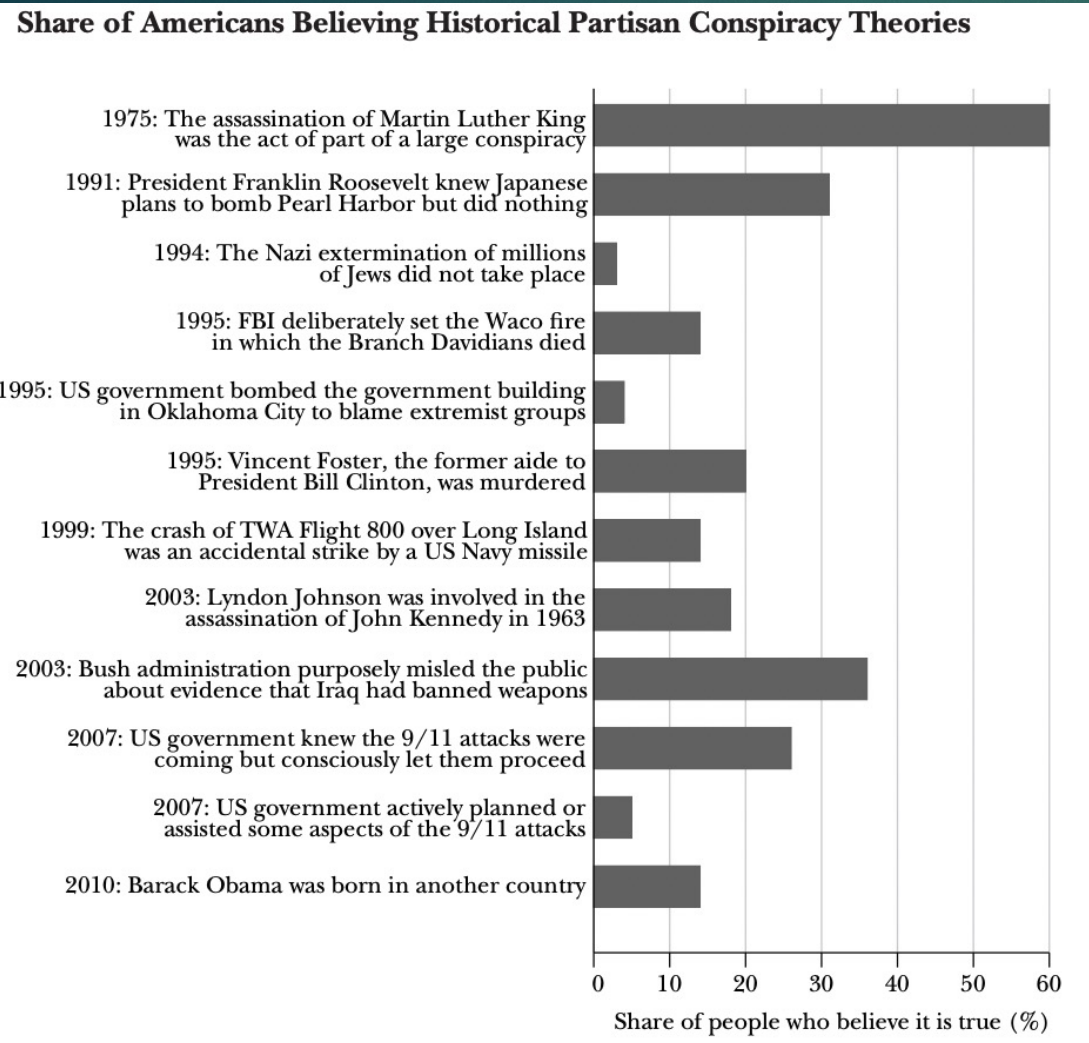
Very mild conditions
for learning



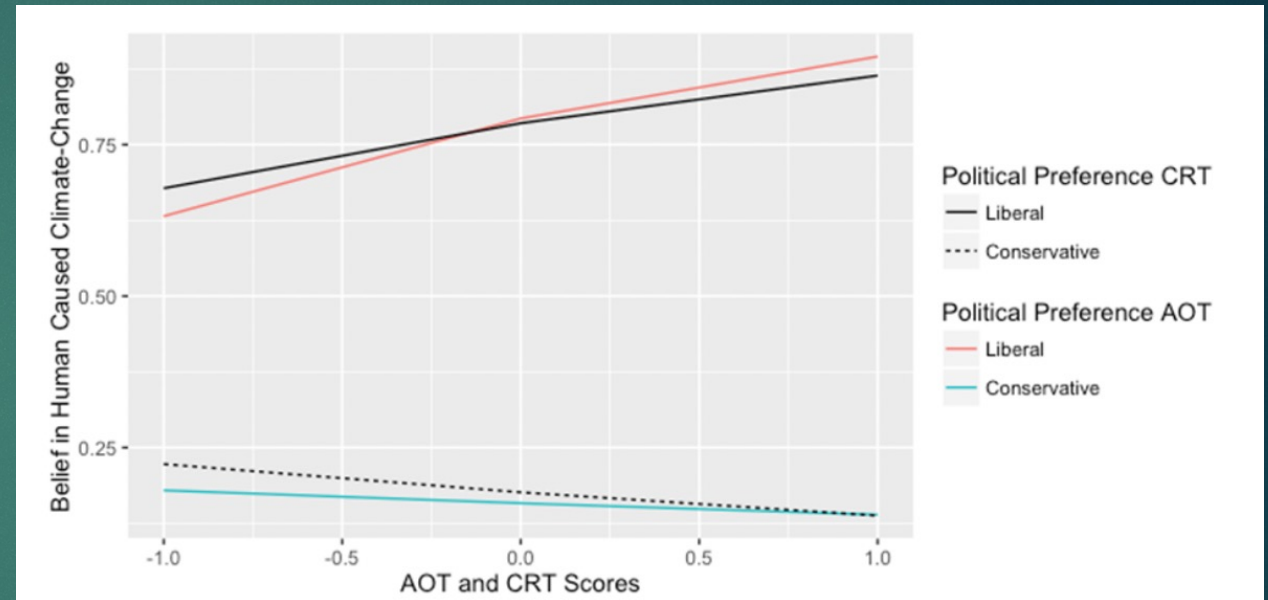
[Unsophisticated]
e.g., Golub et al (2010): Naïve Learning in
Social Networks and the Wisdom of Crowds

Mild (but stricter)
conditions for learning

But there is often mislearning...



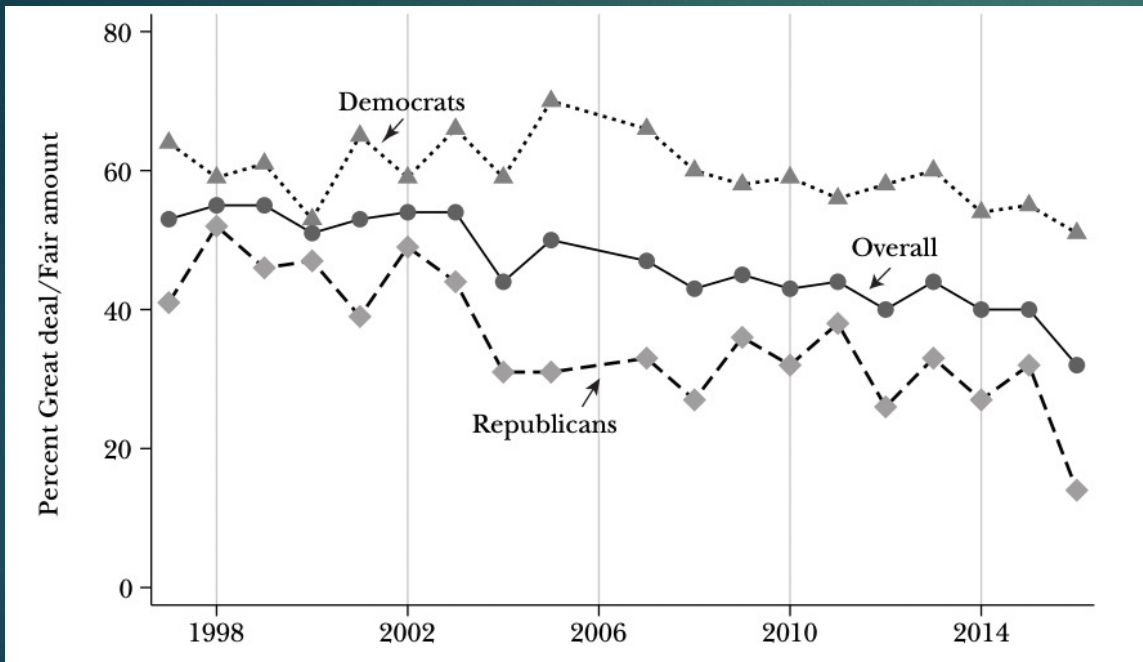
Allcott and Gentzkow (2017)



Higher sophistication can lead to more disagreement on political issues such as climate change

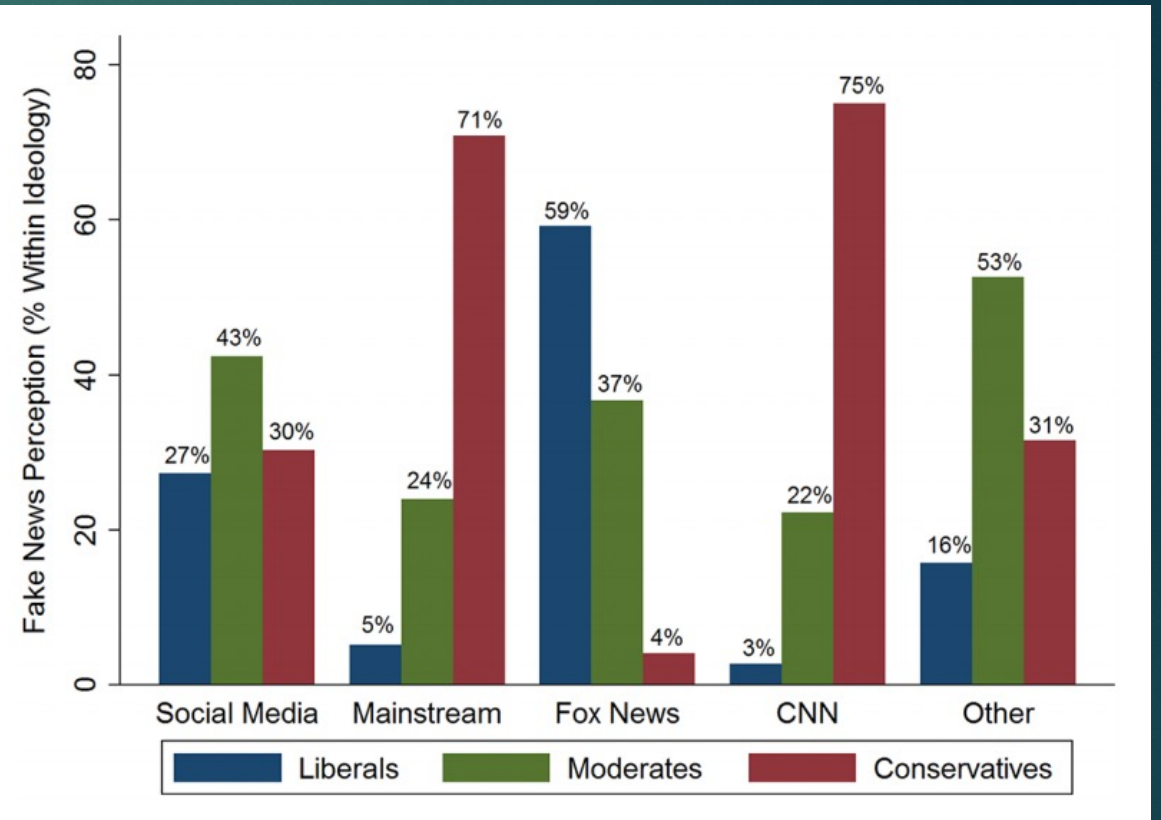
Corbin (2016)

Motivation: Misinformation



Growing distrust in media outlets

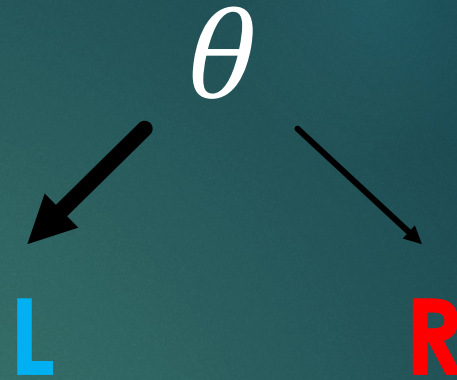
Allcott and Gentzkow (2017)



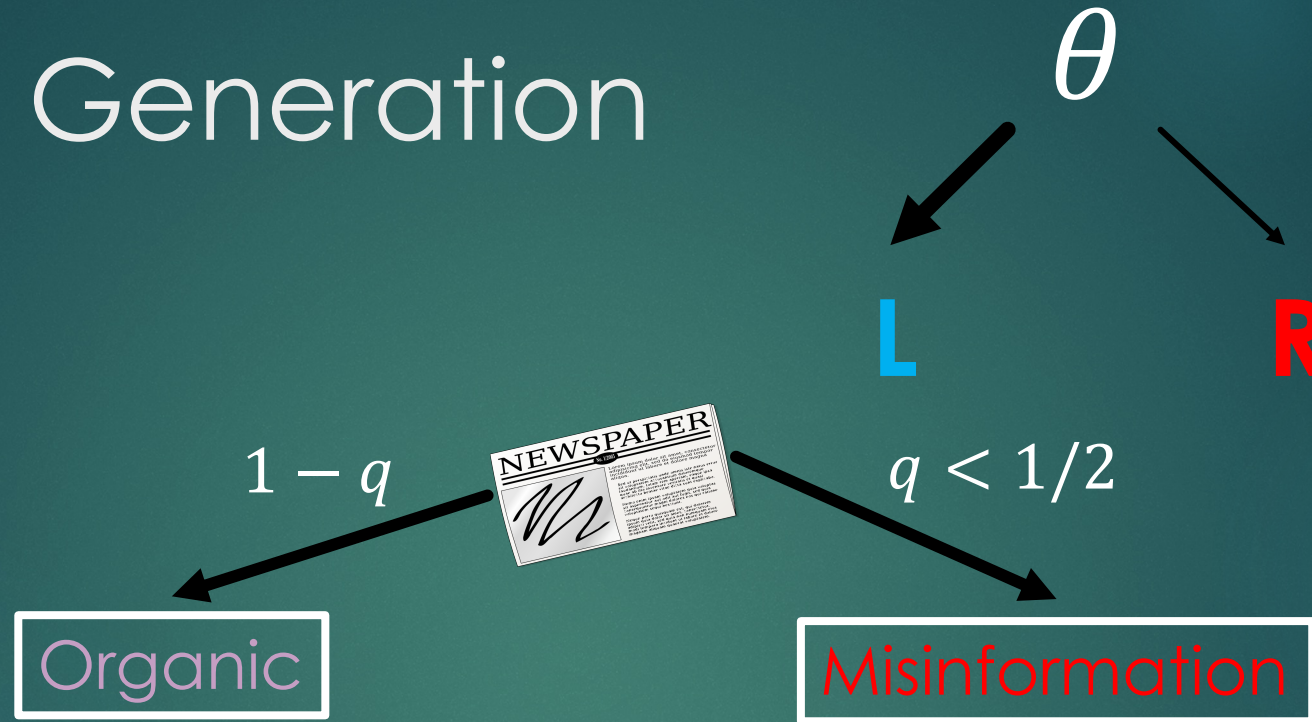
Disagreement over where the misinformation is coming from

van der Linden (2020)

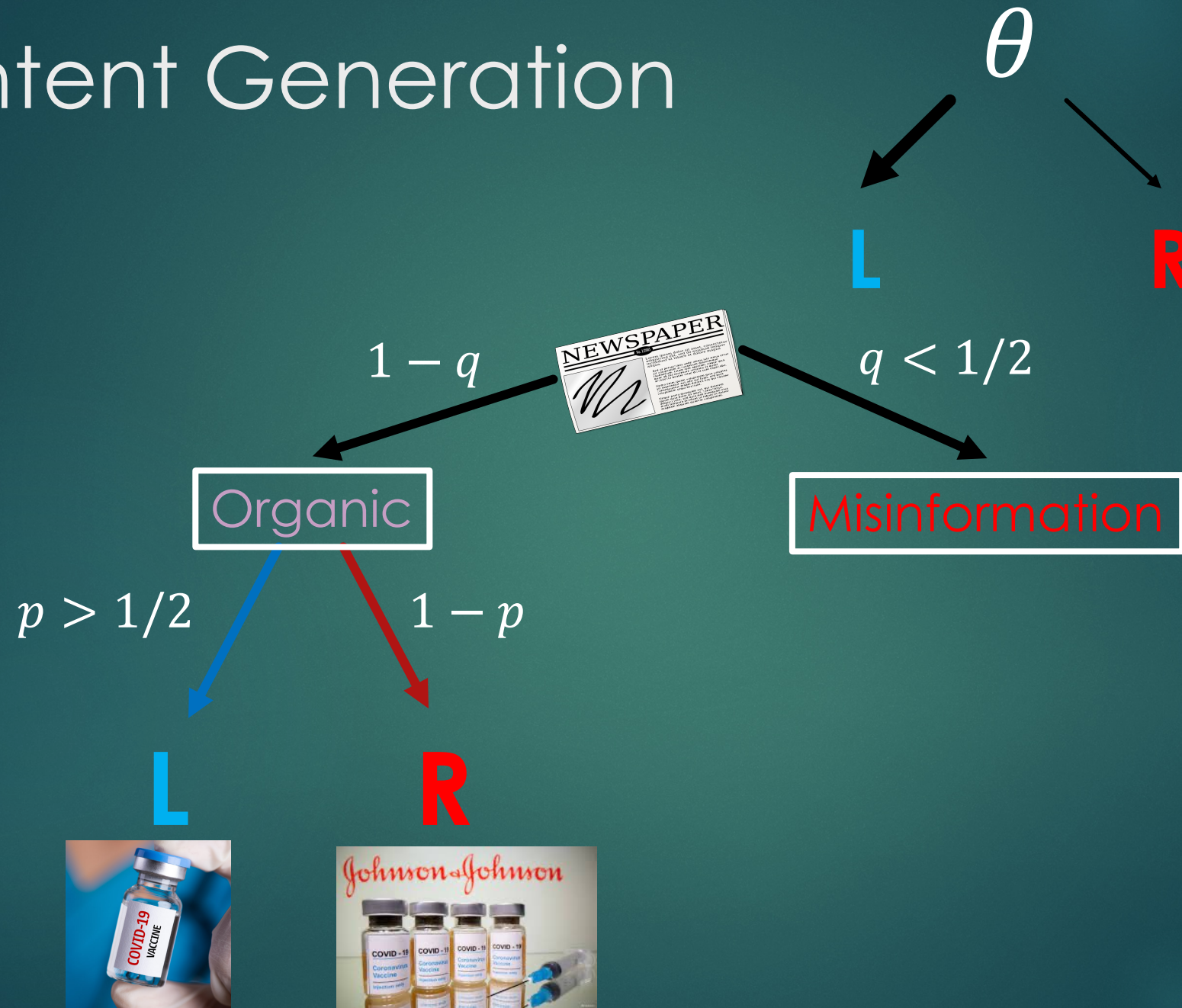
Content Generation



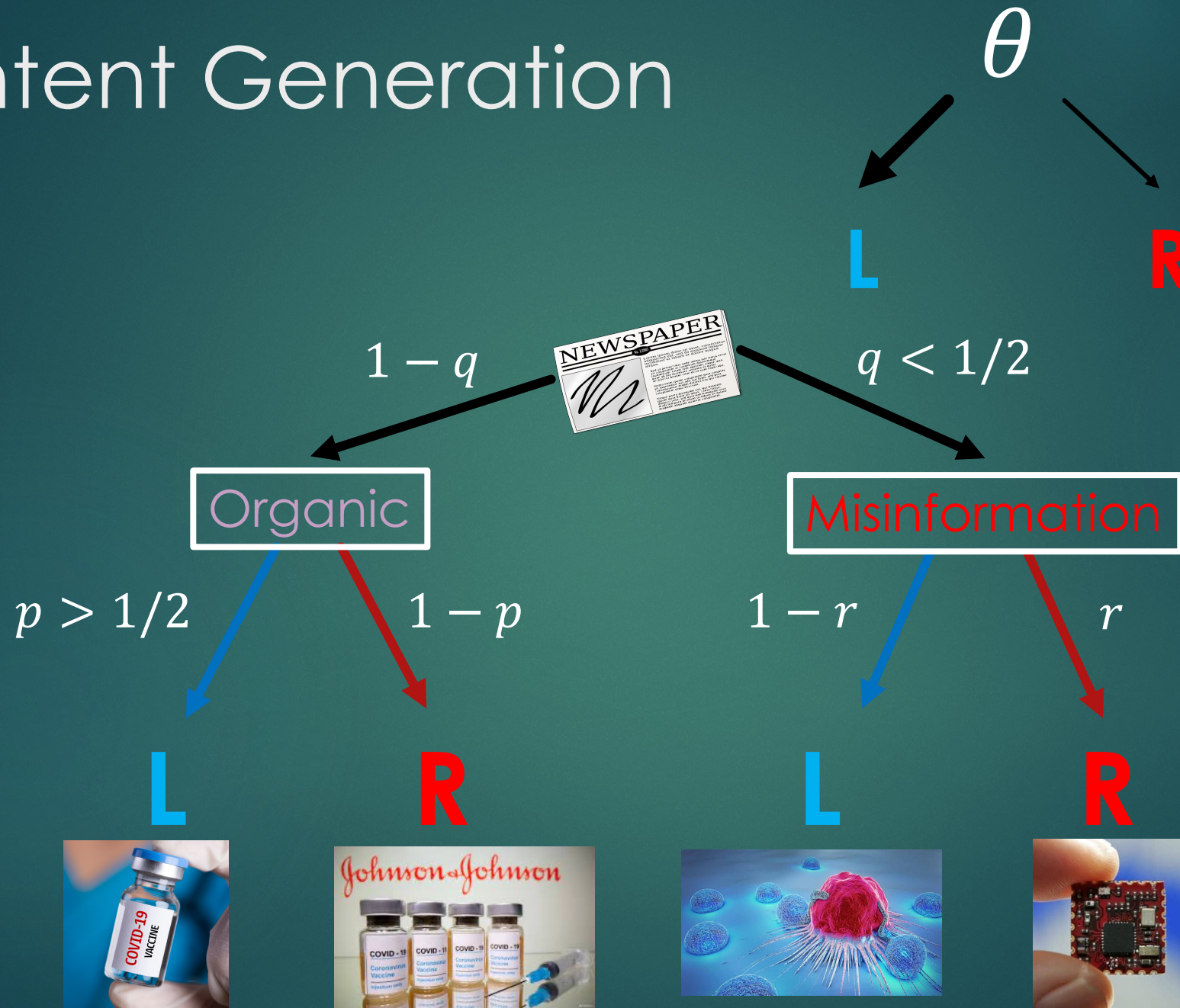
Content Generation



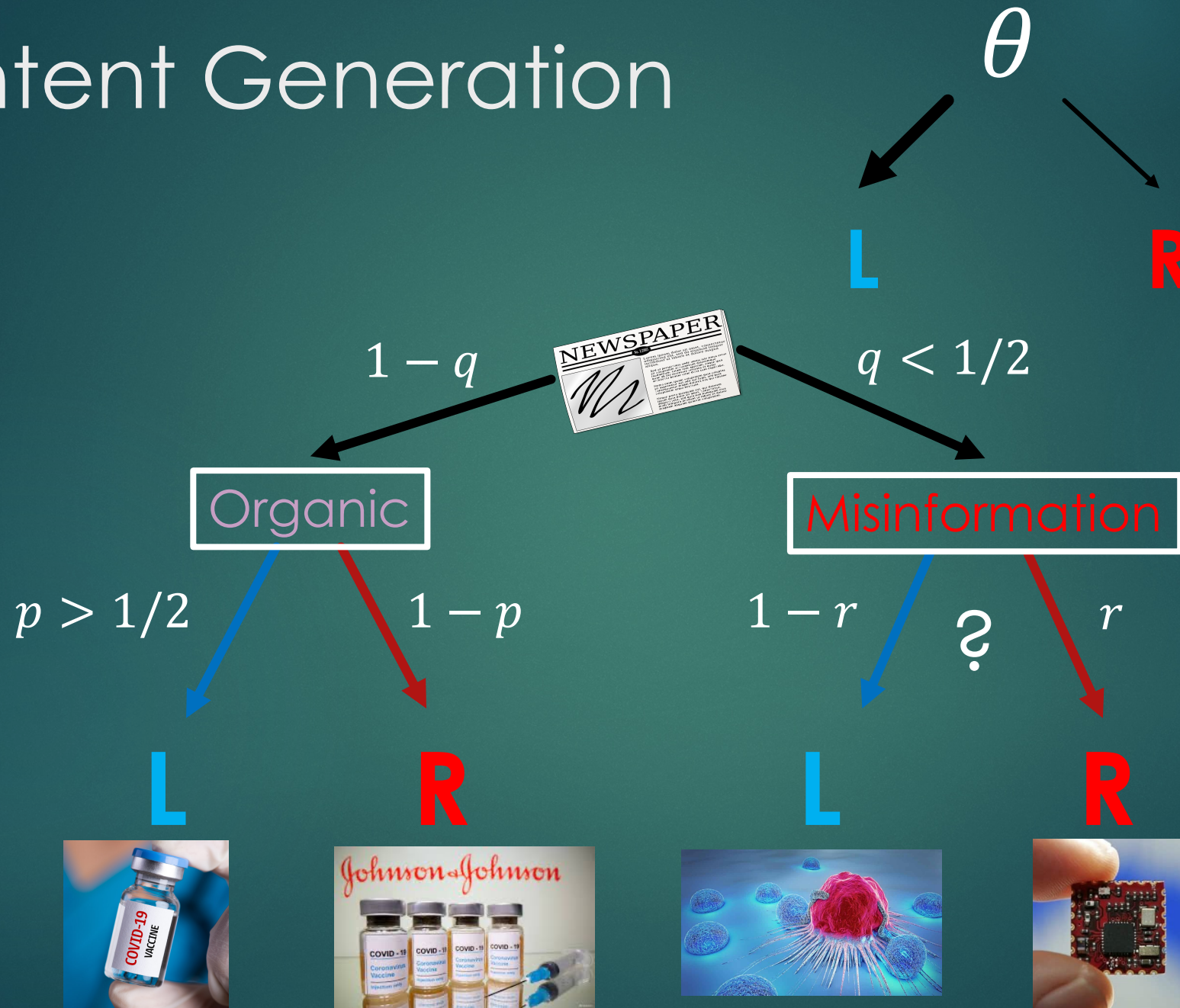
Content Generation



Content Generation

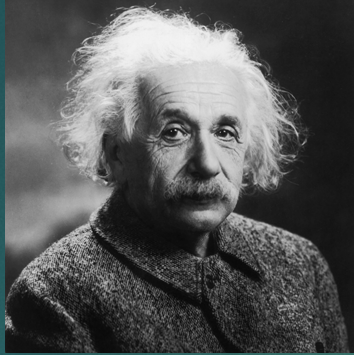


Content Generation



Sophistication Types

Sophisticated (Bayesian)



$$b_i \in [0,1] \sim H$$

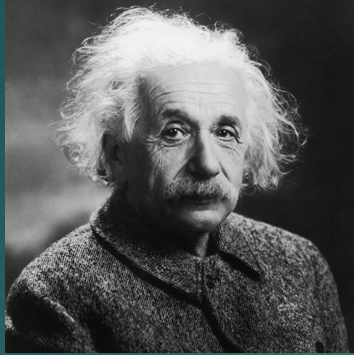
Unsophisticated (DeGroot)



$$b_i \in [0,1] \sim H$$

Sophistication Types

Sophisticated (Bayesian)



$$b_i \in [0,1] \sim H$$



$$m_i = \mathbf{R}$$

Unsophisticated (DeGroot)



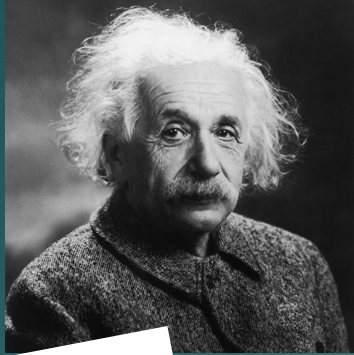
$$b_i \in [0,1] \sim H$$



$$m_i = \mathbf{R}$$

Sophistication Types

Sophisticated (Bayesian)



Signal + Perception of Accuracy

$[0, 1] \sim H$

$m_i = R$

$$\pi_1 = \int_0^1 \frac{p(1-q)b_i + qr}{p(1-q)b_i + (1-p)(1-q)(1-b_i) + qr} f(r) dr$$

Unsophisticated (DeGroot)



$b_i \in [0, 1]$

Face Value of Signal

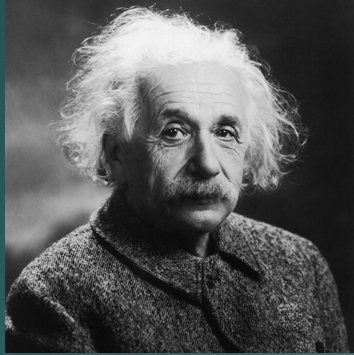
$m_i = R$

pb_i

$$\pi_1 = \frac{pb_i}{pb_i + (1-p)(1-b_i)}$$

Sophistication Types

Sophisticated (Bayesian)

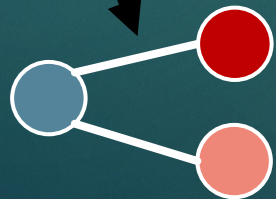


$$b_i \in [0,1] \sim H$$

$$m_i = R$$

Ignore Redundant Information from Others

$$\pi_i = \int_0^1 \frac{b_i + qr}{p(1-q)(1-b_i) + qr} f(r) dr$$



Network Inference

Unsophisticated (DeGroot)

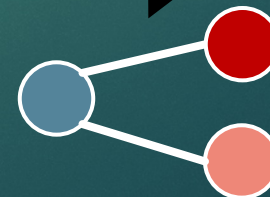


$$b_i \in [0,1] \sim H$$

$$m_i = R$$

Rule-of-Thumb Learning

$$\pi_i = \frac{pb_i}{pb_i + (1-b_i)}$$



Average Other Beliefs with Yours

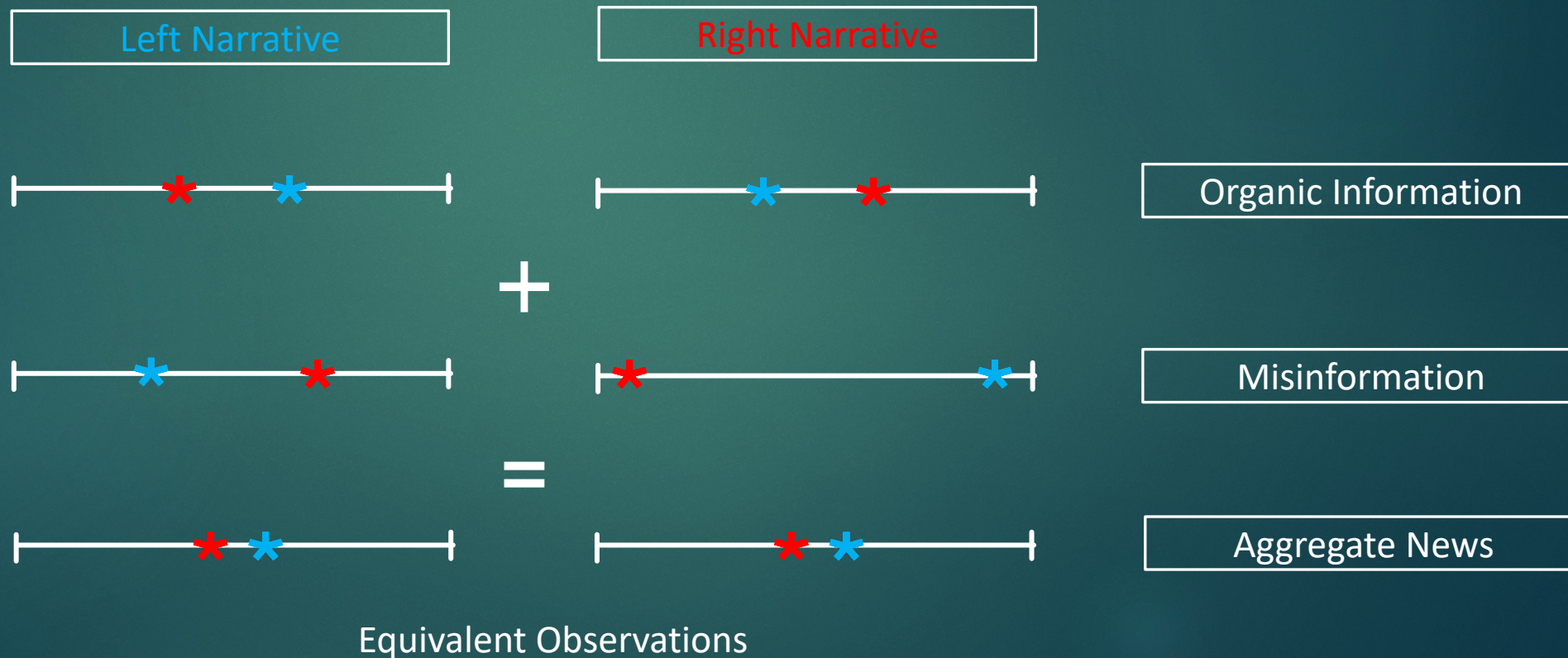
What breaks learning?

Unsophisticated: Learning occurs if and only if misinformation does not advocate too much for the opposite of θ (i.e., $r < r_D^*$ for some r_D^*).

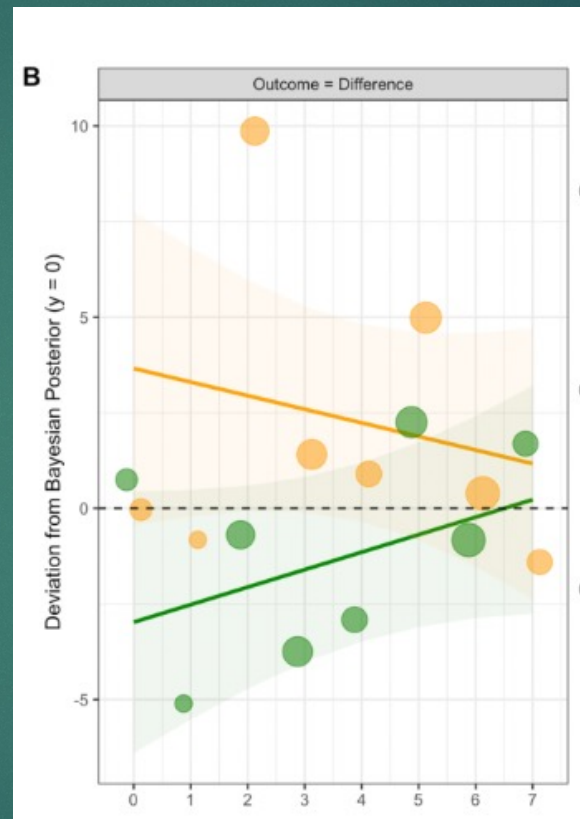
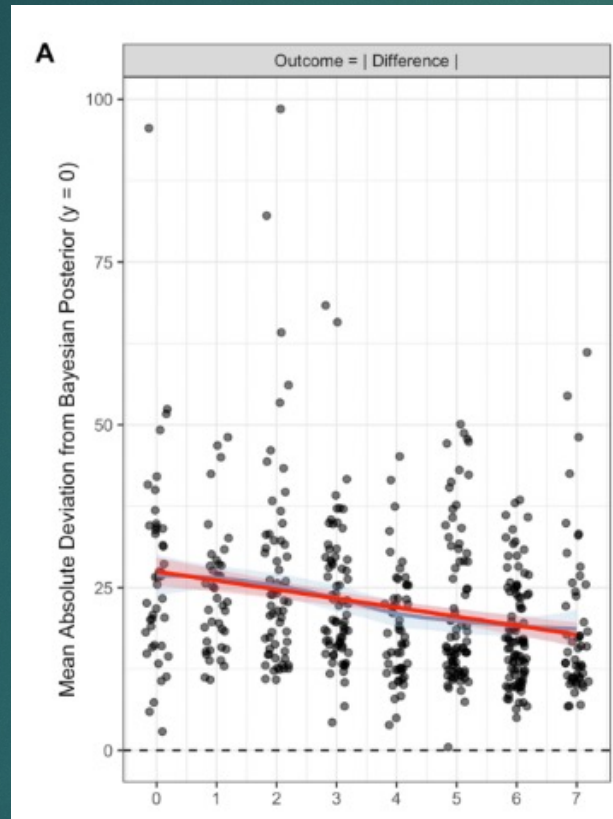
What breaks learning?

Unsophisticated: Learning occurs if and only if misinformation does not advocate too much for the opposite of θ (i.e., $r < r_D^*$ for some r_D^*).

Sophisticated: Learning occurs if there is a unique narrative.



Biased or Bayesian updating?



CRT Sum Score

Tappin, Pennycook, and Rand (2019)

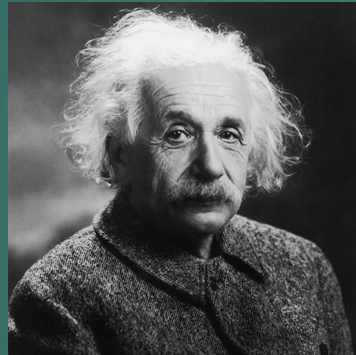
Main Characterization

Who learns better with **organic** information? **Some** misinformation? **Mostly** misinformation?

Main Characterization

Who learns better with **organic** information? **Some** misinformation? **Mostly** misinformation?

Low q



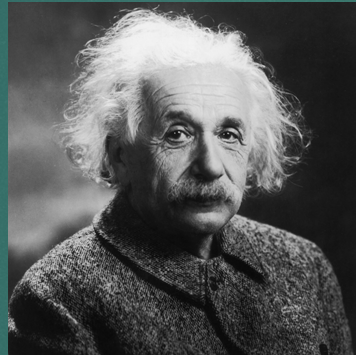
>



Main Characterization

Who learns better with **organic** information? **Some** misinformation? **Mostly** misinformation?

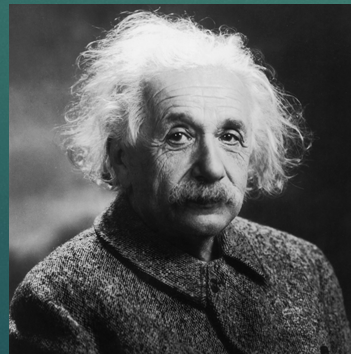
Low q



>




High q



<





**How does misinformation
regulation affect the learning of
different sophistication types?**



Policy 1: Diverse Content

Policy 2: Censorship

Policy 3: Accuracy Nudging

Policy 4: Performance Targets

Policy 1: Diverse Content Provision

- ▶ FCC Fairness Doctrine 1949; repealed in 1987 led to one-sided coverage.

Policy 1: Diverse Content Provision

- ▶ FCC Fairness Doctrine 1949; repealed in 1987 led to one-sided coverage.
- ▶ Increase the likelihood that the message distribution presents more evenly distributed content.
 - ▶ **Diverse content provision:** if many **L** articles shown in the past, increase likelihood of showing **R**.

Policy 1: Diverse Content Provision

- ▶ FCC Fairness Doctrine 1949; repealed in 1987 led to one-sided coverage.
- ▶ Increase the likelihood that the message distribution presents more evenly distributed content.
 - ▶ **Diverse content provision:** if many **L** articles shown in the past, increase likelihood of showing **R**.
- ▶ *Benefit:* Reduces the likelihood that misinformation is heavily skewed toward either **L** or **R**.
- ▶ *Potential Cost:* Reduces the strength of the organic content.

Policy 1: Diverse Content Provision

- ▶ FCC Fairness Doctrine 1949; repealed in 1987 led to one-sided coverage.
- ▶ Increase the likelihood that the message distribution presents more evenly distributed content.
 - ▶ **Diverse content provision:** if many **L** articles shown in the past, increase likelihood of showing **R**.
- ▶ *Benefit:* Reduces the likelihood that misinformation is heavily skewed toward either **L** or **R**.
- ▶ *Potential Cost:* Reduces the strength of the organic content.
- ▶ Similar conclusions hold for “counter-attitudinal” news where content is provided intentionally against belief.

Policy 1: Diverse Content Provision

- ▶ Unsophisticated: Always effective and ultimately allows organic (but weaker) information to dominate.

Policy 1: Diverse Content Provision

- ▶ Unsophisticated: Always effective and ultimately allows organic (but weaker) information to dominate.
- ▶ Sophisticated: “Pandora’s box” effect. Effective provided q is not too large.
 - ▶ When q is **small**, sophisticated agents can thrive. Pulling content toward center prevents latching onto separate narratives.
 - ▶ When q is **large**, sophisticated agents do worse. Pulling content toward center permits the telling of more drastic narratives.

Policy 1: Diverse Content Provision

- ▶ Unsophisticated: Always effective and ultimately allows organic (but weaker) information to dominate.
- ▶ Sophisticated: “Pandora’s box” effect. Effective provided q is not too large.
 - ▶ When q is **small**, sophisticated agents can thrive. Pulling content toward center prevents latching onto separate narratives.
 - ▶ When q is **large**, sophisticated agents do worse. Pulling content toward center permits the telling of more drastic narratives.
- ▶ Counter-attitudinal content induces more sympathy: [Levy \(2021\)](#)
- ▶ Counter-attitudinal leads to more rejection of other side: [Bail et al \(2018\)](#)

Policy 2: Censorship

Facebook Takes Down Viral Video Making False Claim That 'Hydroxychloroquine Cures Covid'

This Tweet violated the Twitter Rules about [specific rule]. However, Twitter has determined that it may be in the public's interest for the Tweet to remain accessible. [Learn more](#)

Policy 2: Censorship

Facebook Takes Down Viral Video Making False Claim That 'Hydroxychloroquine Cures Covid'

This Tweet violated the Twitter Rules about [specific rule]. However, Twitter has determined that it may be in the public's interest for the Tweet to remain accessible. [Learn more](#)

"Many governments are grappling with how to approach the spread of misinformation, but few have outlawed it. As the UN and others have noted, the general criminalization of sharing misinformation would be 'incompatible with international standards for restrictions on freedom of expression.'"

Facebook White Paper *Charting A Way Forward: Online Content Regulation*

Policy 2: Censorship

Facebook Takes Down Viral Video Making False Claim That 'Hydroxychloroquine Cures Covid'

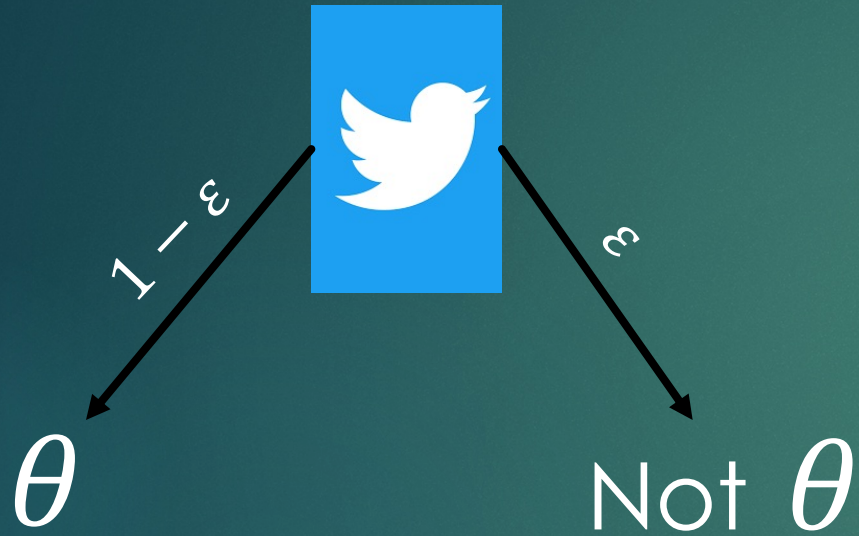
This Tweet violated the Twitter Rules about [specific rule]. However, Twitter has determined that it may be in the public's interest for the Tweet to remain accessible. [Learn more](#)

"Many governments are grappling with how to approach the spread of misinformation, but few have outlawed it. As the UN and others have noted, the general criminalization of sharing misinformation would be 'incompatible with international standards for restrictions on freedom of expression.'"

Facebook White Paper *Charting A Way Forward: Online Content Regulation*

Question: Do restrictions on freedom of expression (when removing content that contains misinformation) ever hurt learning?

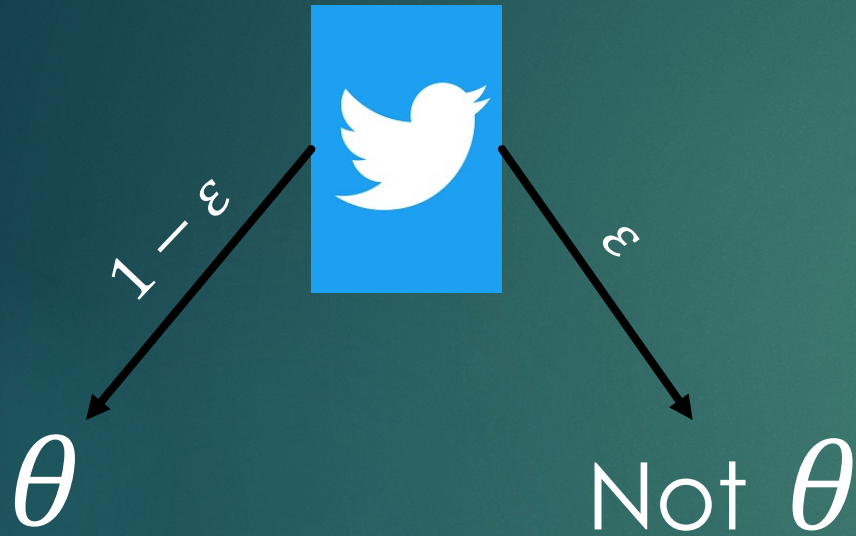
Policy 2: Censorship



Research

ϵ is assumed to be small**

Policy 2: Censorship



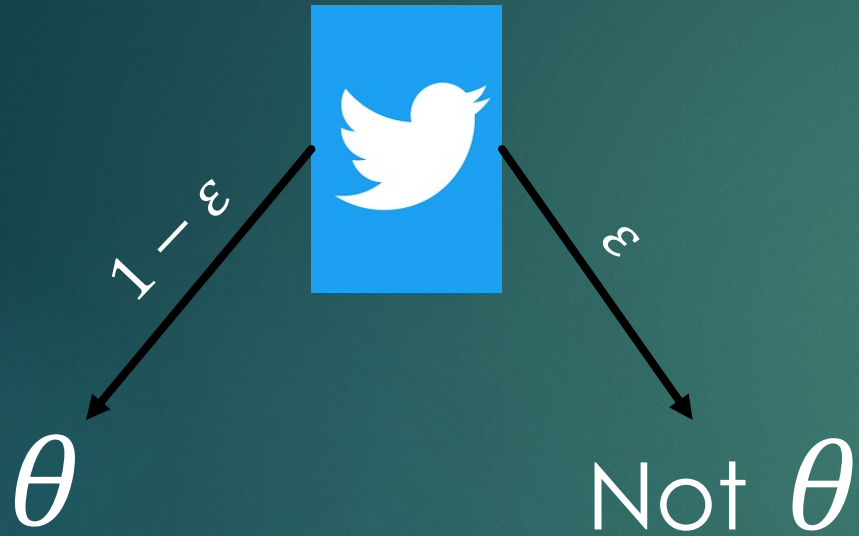
Research

ϵ is assumed to be small**



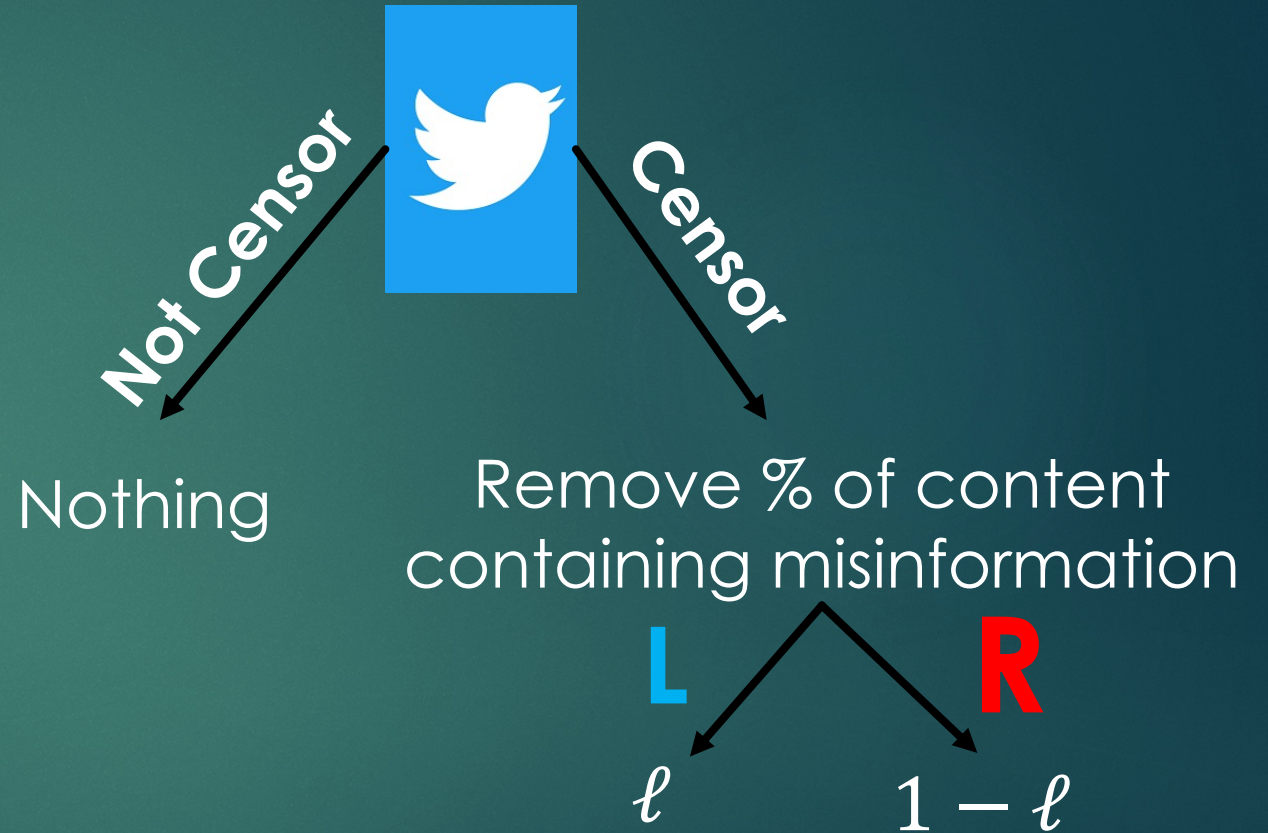
Allow Full Freedom of Expression?

Policy 2: Censorship



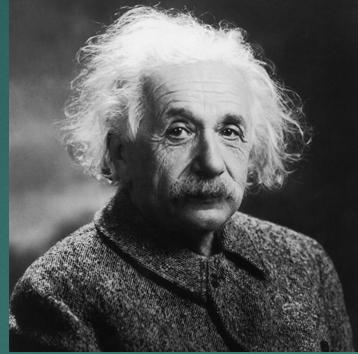
Research

ϵ is assumed to be small**



Allow Full Freedom of Expression?

Policy 2: Censorship



Learning stays the same

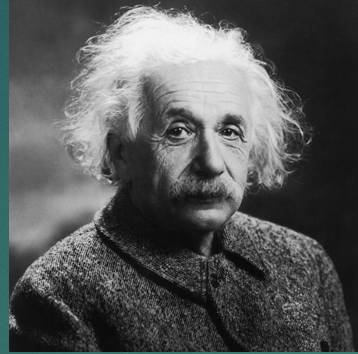


Learning stays the same

Not Censor



Policy 2: Censorship



Learning stays the same



Learning stays the same

Wilder **narratives** possible – believe the platform is part of the misinformation problem.
→ Learning is worse

Remove all **R** content
→ Learning improves


Censor Not Censor



Policy 3: Accuracy Nudging

- ▶ As suggested by Pennycook et al (2020); Pennycook et al (2021): nudge platform users to think critically about the presence of misinformation.

HOW ACCURATE IS THIS HEADLINE?

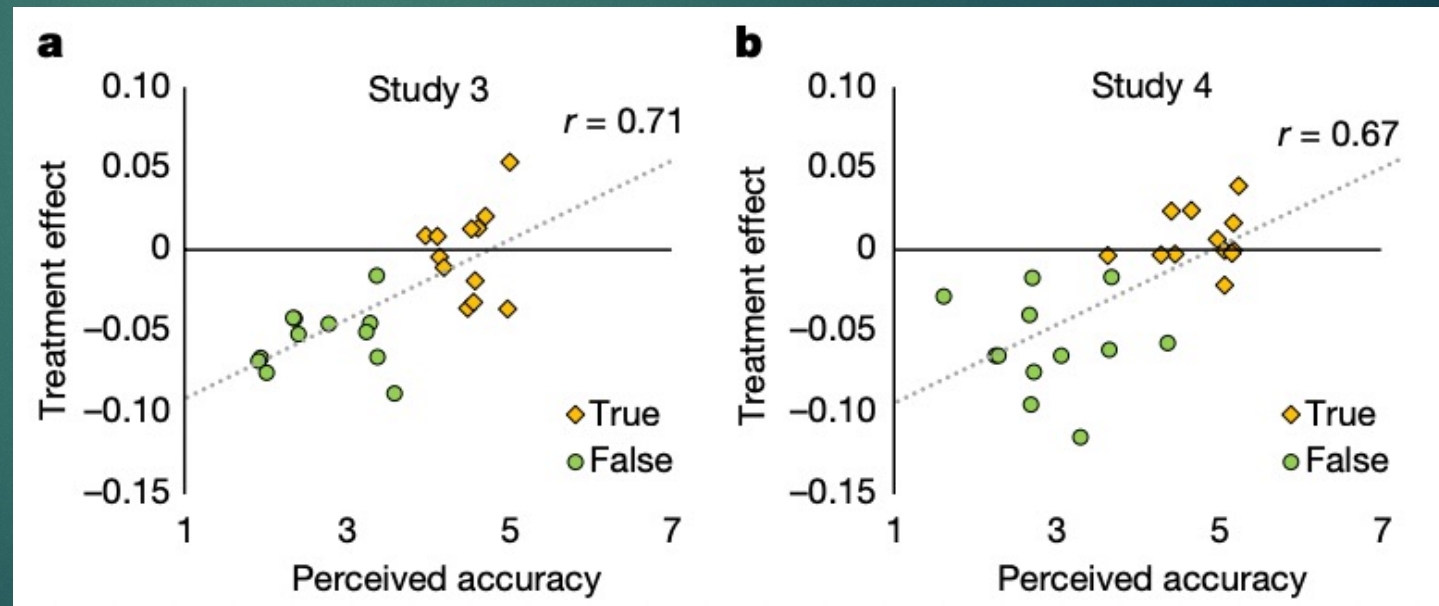


Woman who had ovary frozen in childhood give...
She is believed to be the first woman in the world to have a baby after having ovarian tissue frozen befo...
surveycamel.com

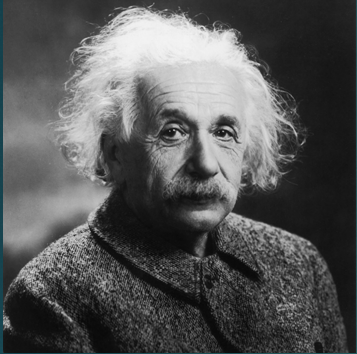
Thanks for following me! Can I ask you a favor? I'm wondering how accurate the above headline is, and I'm doing a survey to find out.
surveycamel.com/ze/news/story5...

Based on the headline, do you think it is accurately describing something that actually happened?

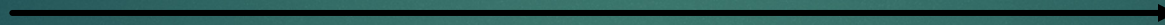
Please rate as: 1=Not at all accurate, 2 = Not very accurate, 3= Somewhat accurate, 4 = Very accurate



Policy 3: Accuracy Nudging

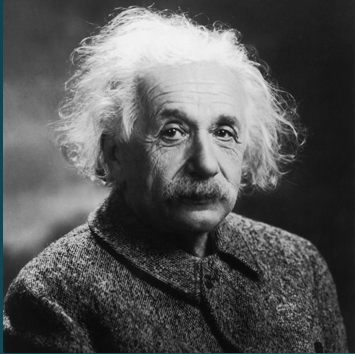


Already fully Bayesian inference

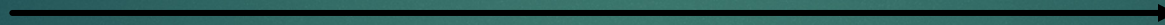


Does not help or hurt learning

Policy 3: Accuracy Nudging



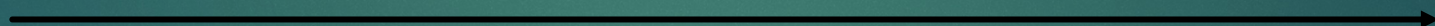
Already fully Bayesian inference



Does not help or hurt learning



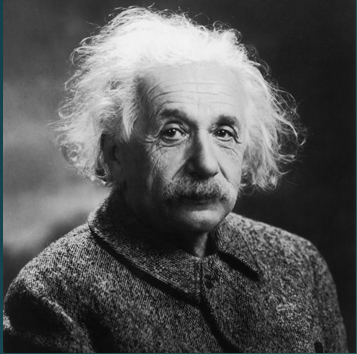
Small fraction of agents update on perception of accuracy:



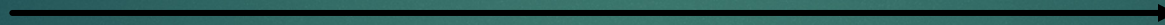
$$\pi_1 = \int_0^1 \frac{p(1-q)b_i + qr}{p(1-q)b_i + (1-p)(1-q)(1-b_i) + qr} f(r) dr$$

Helps learning

Policy 3: Accuracy Nudging



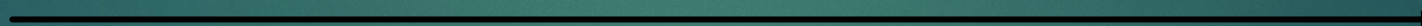
Already fully Bayesian inference



Does not help or hurt learning



Small fraction of agents update on perception of accuracy:



$$\pi_1 = \int_0^1 \frac{p(1-q)b_i + qr}{p(1-q)b_i + (1-p)(1-q)(1-b_i) + qr} f(r) dr$$

Helps learning

- ▶ In high-misinformation environments, accuracy nudged unsophisticated agents are the most resistant to misinformation.

Policy 4: Performance Targets



“Governments could also consider requiring companies to hit specific performance targets, such as decreasing the prevalence of content...[with] policy violations. While such targets may have benefits, they could also create perverse incentives for companies to find ways to decrease enforcement burdens.”

Facebook White Paper *Charting A Way Forward: Online Content Regulation*

Policy 4: Performance Targets

“Governments could also consider requiring companies to hit specific performance targets, such as decreasing the prevalence of content...[with] policy violations. While such targets may have benefits, they could also create perverse incentives for companies to find ways to decrease enforcement burdens.”

Facebook White Paper *Charting A Way Forward: Online Content Regulation*

- ▶ Implement a performance target to decrease misinformation.
- ▶ **Ultimate goal:** Reduce the likelihood of mislearning (because of misinformation) to some level $\phi^* > 0$.

Policy 4: Performance Targets

“Governments could also consider requiring companies to hit specific performance targets, such as decreasing the prevalence of content...[with] policy violations. While such targets may have benefits, they could also create perverse incentives for companies to find ways to decrease enforcement burdens.”

Facebook White Paper *Charting A Way Forward: Online Content Regulation*

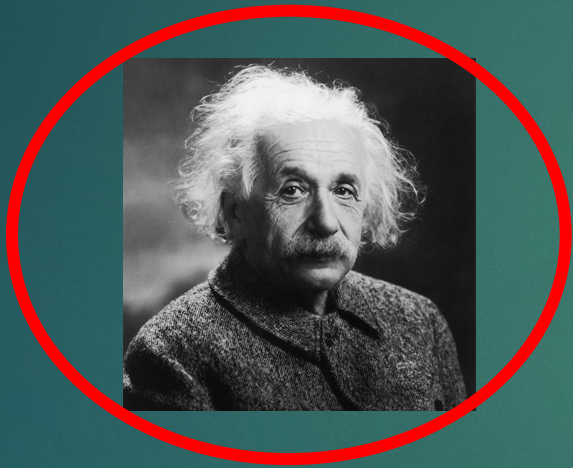
- ▶ Implement a performance target to decrease misinformation.
- ▶ **Ultimate goal:** Reduce the likelihood of mislearning (because of misinformation) to some level $\phi^* > 0$.
- ▶ There is a *moral hazard* cost to decreasing the target more.
 - ▶ Define misinformation more narrowly.
 - ▶ Make reporting of misinformation more difficult.
 - ▶ Reduce efforts to stop misinformation that is already viral.

Policy 4: Performance Targets

Once misinformation becomes a problem, which type of population should be regulated more?

Policy 4: Performance Targets

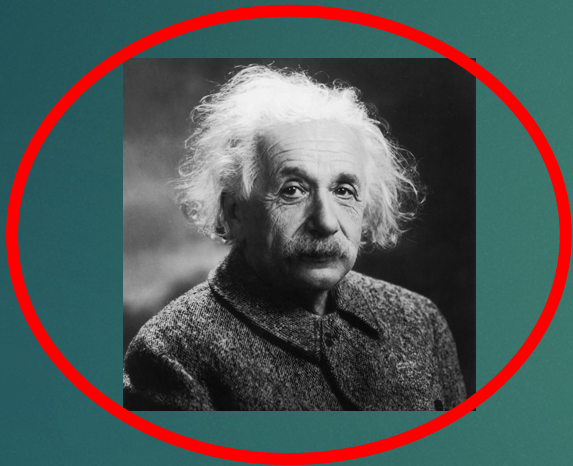
Once misinformation becomes a problem, which type of population should be regulated more?



- ▶ Need to set a low performance target for sophisticated agents to mitigate the ability to dismiss other perspectives as misinformation.

Policy 4: Performance Targets

Once misinformation becomes a problem, which type of population should be regulated more?



- ▶ Need to set a low performance target for sophisticated agents to mitigate the ability to dismiss other perspectives as misinformation.
- ▶ Can get away with setting **less stringent** targets with unsophisticated agents.
- ▶ Regulating the wrong type of population can backfire.

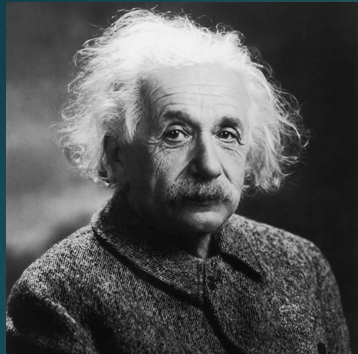
Summary of Policies

Diverse Content

Censorship

Accuracy Nudging

Performance Targets



Could backfire

Backfire

Neutral

Could backfire



Effective

Effective

Effective

Could backfire

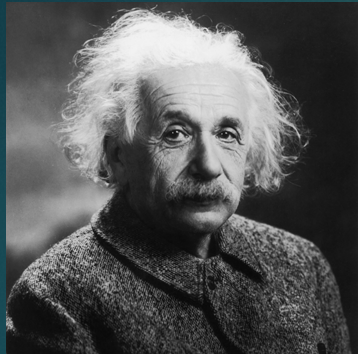
Combining Policies?

Diverse Content

Censorship

Accuracy Nudging

Performance Targets



Could backfire

Backfire

Neutral

Could backfire



Effective

Effective

Effective

Could backfire

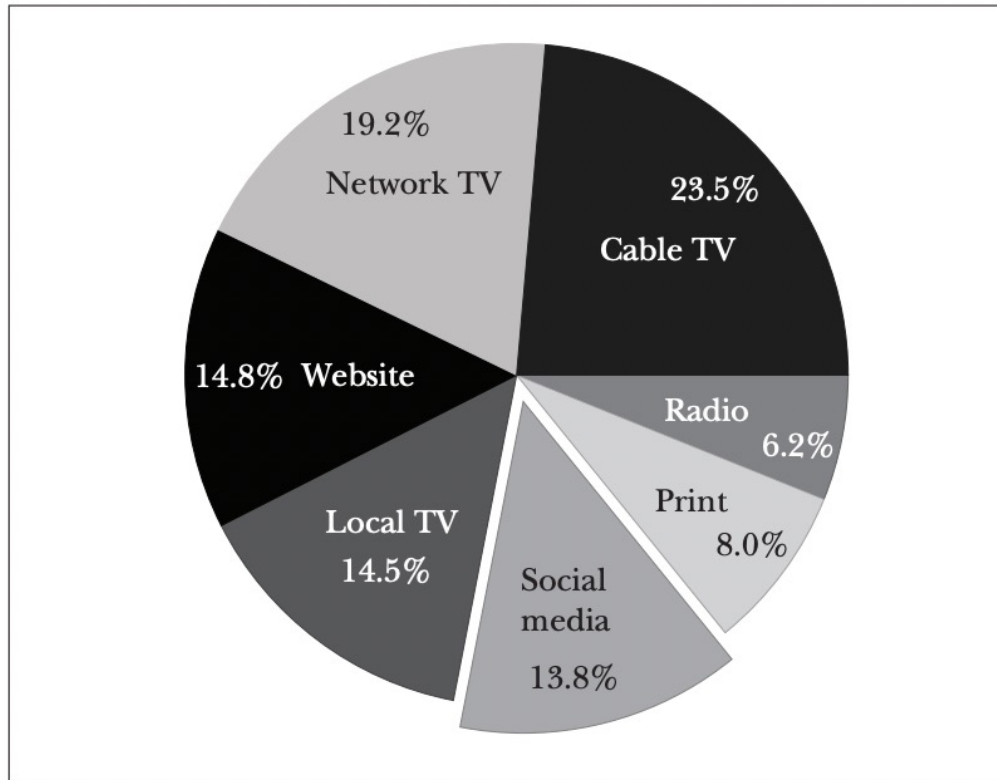


Misinformation: Strategic Sharing, Homophily, and Endogenous Echo Chambers

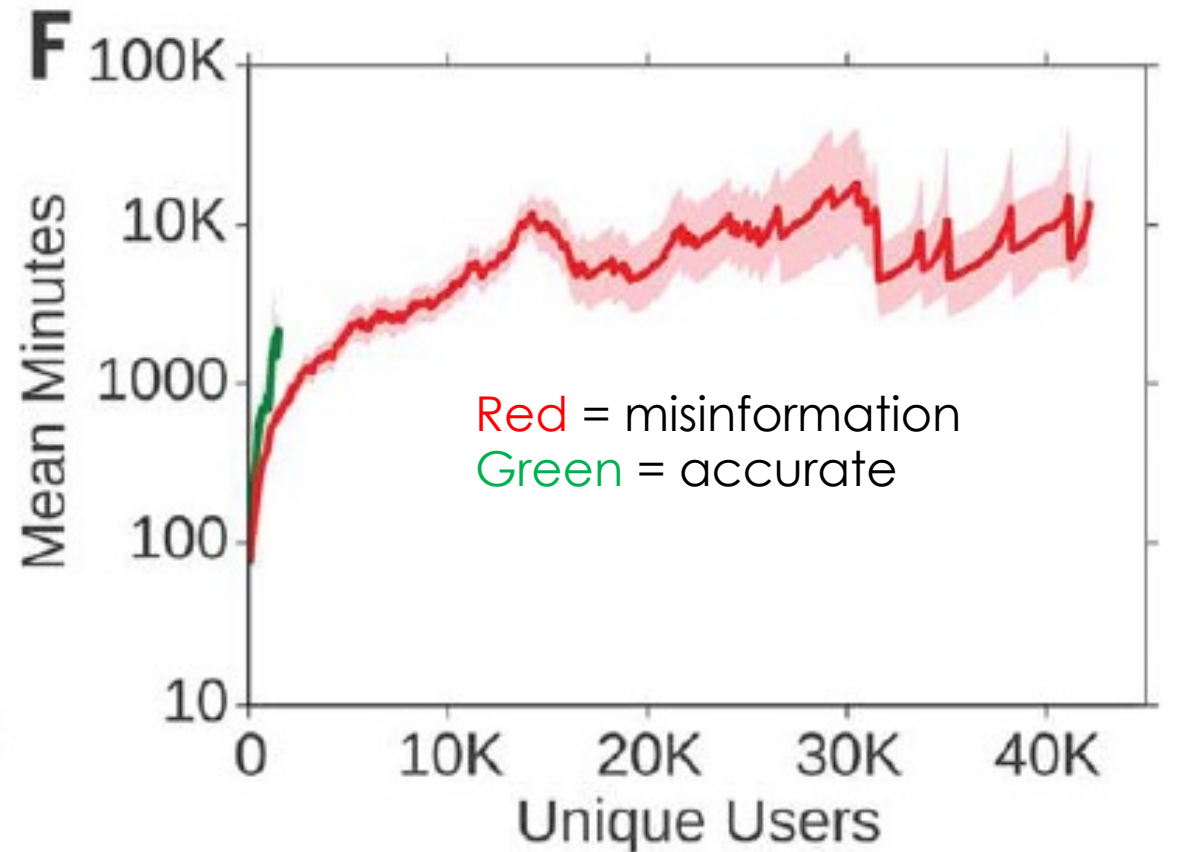
MODEL OF SHARING

Motivation: Platform Sharing

Most Important Source of 2016 Election News

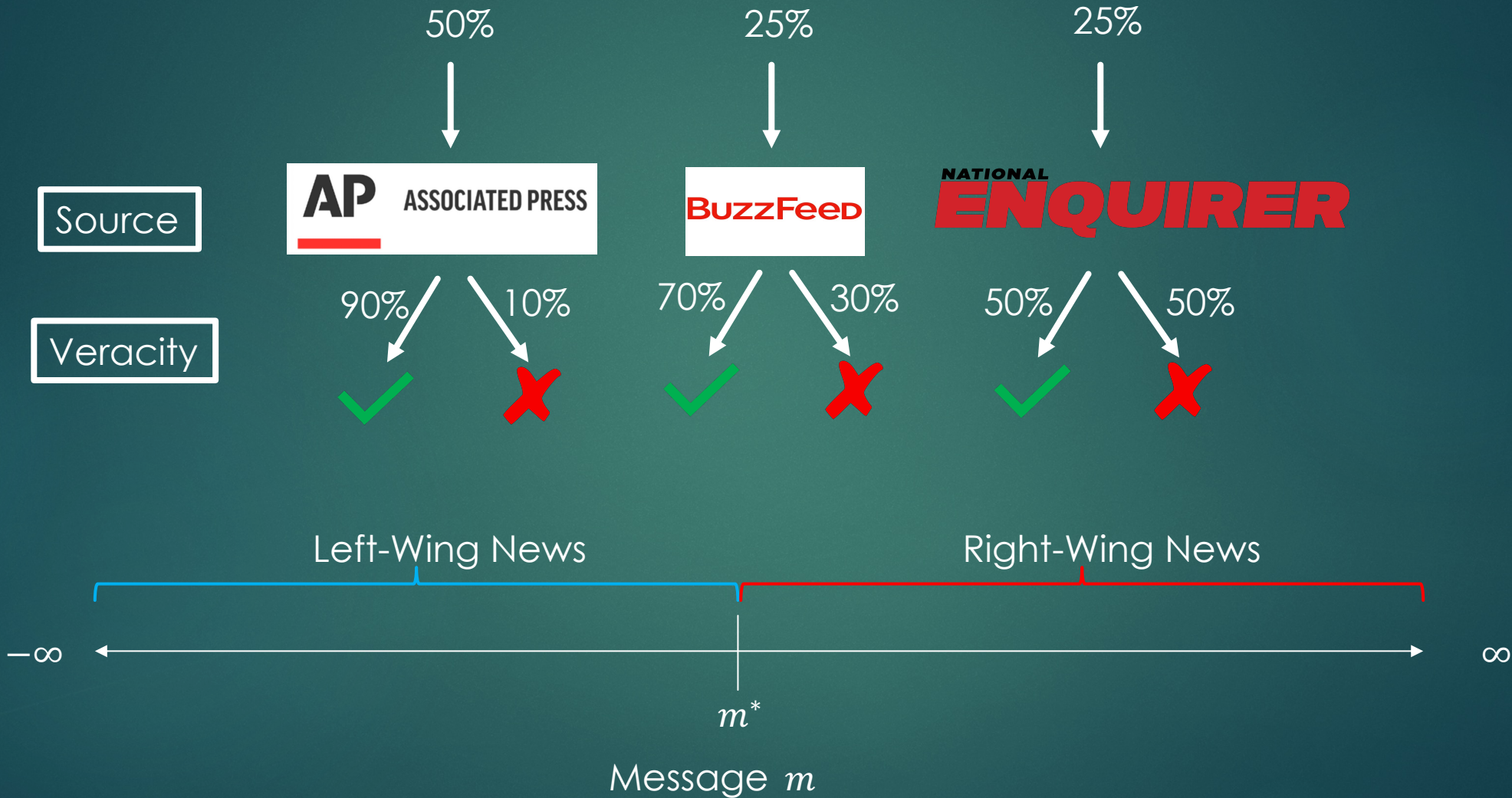


Allcott and Gentzkow (2017)

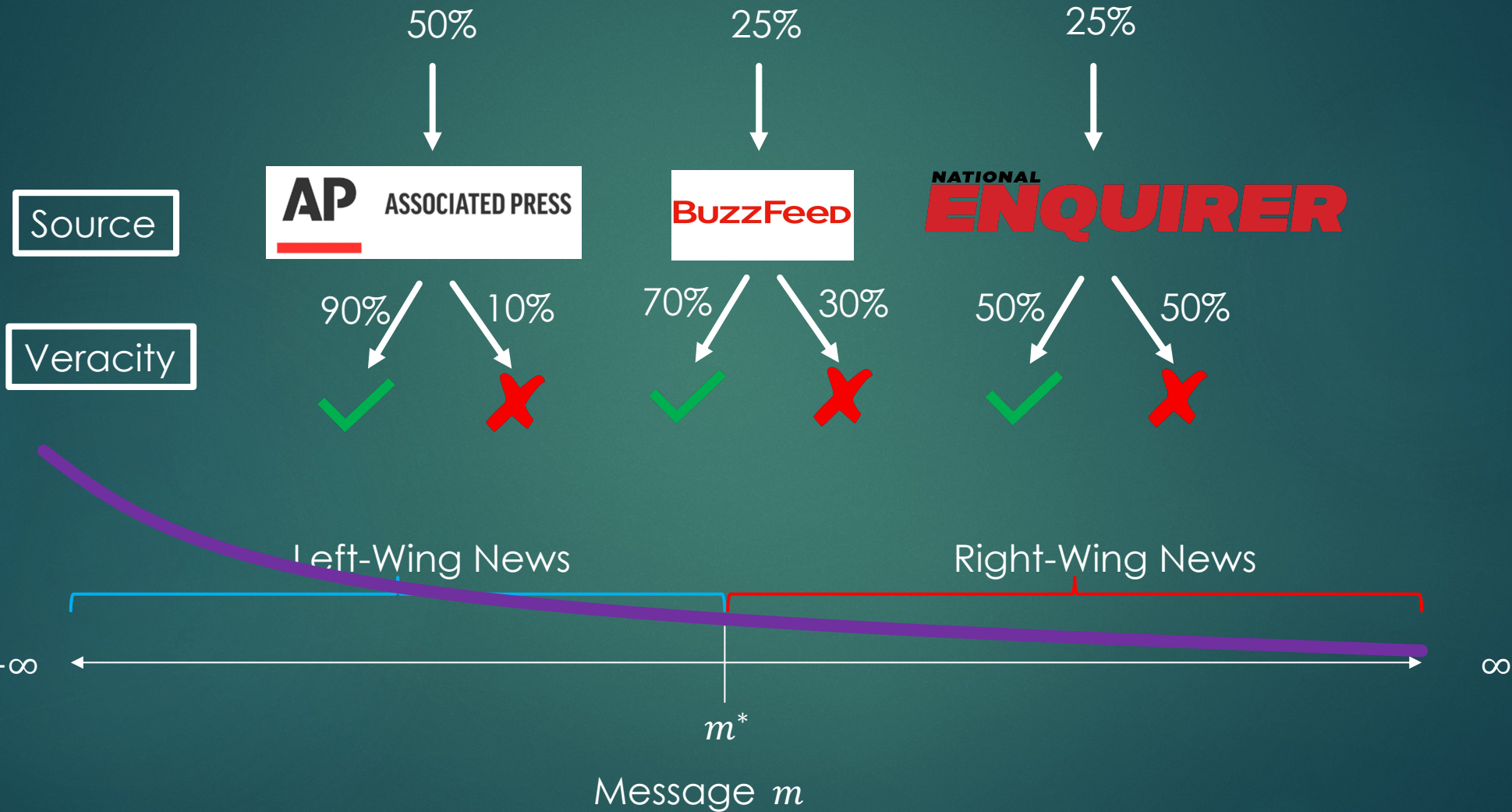


Vosoughi et al (2018)

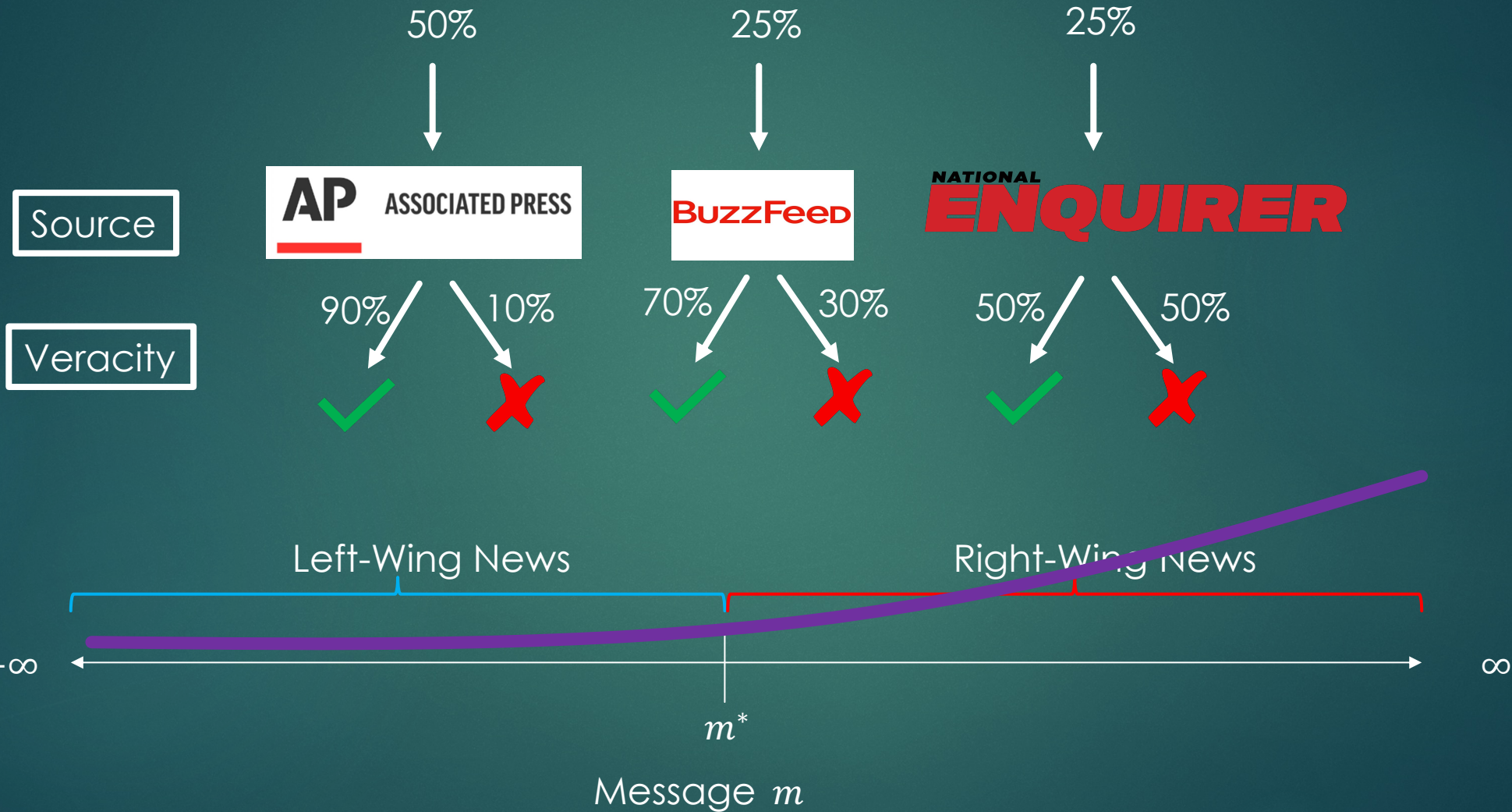
Model: News Generation



Model: News Generation



Model: News Generation

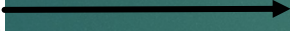


Model: Agents' Actions

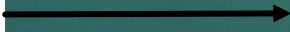
Share



Inspect



Kill



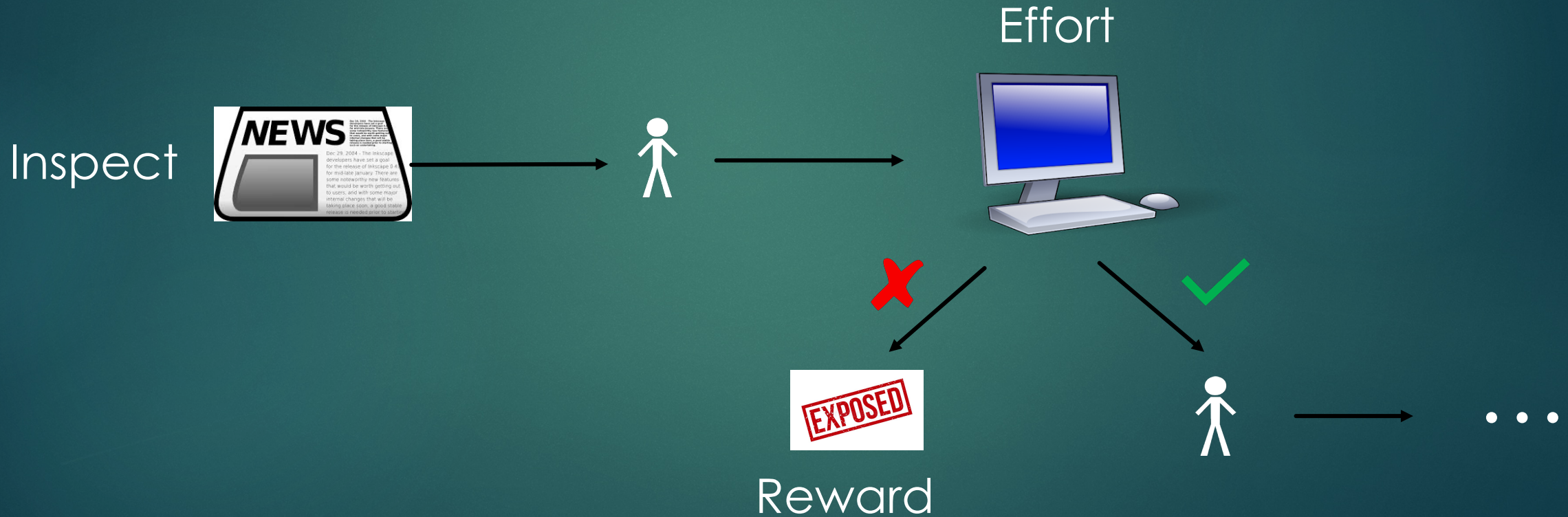
Model: Kill Payoff

Kill



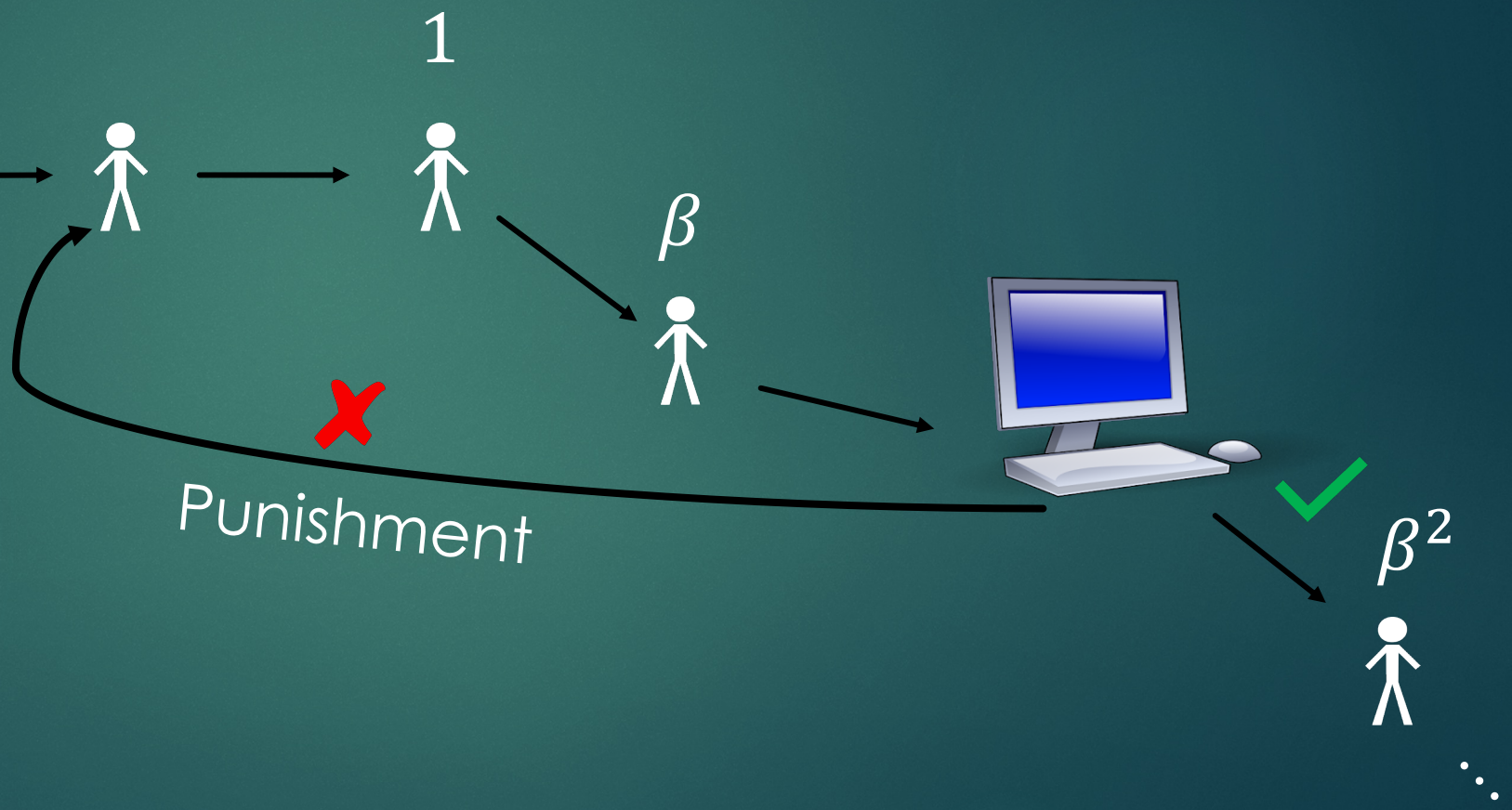
0

Model: Inspect Payoff



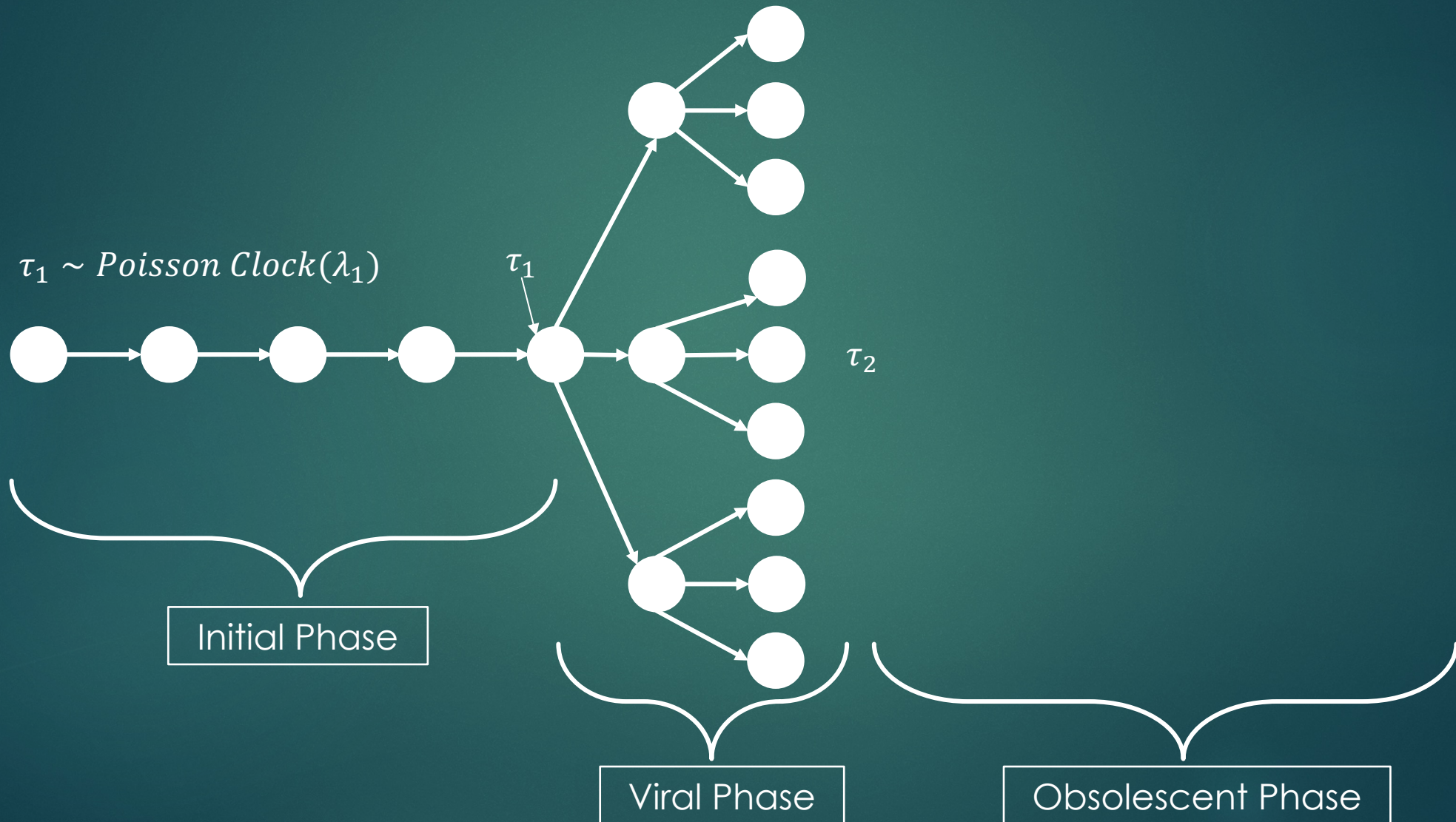
Model: Share Payoff

Share



Model: Lifetime of the Article

$$\tau_2 \sim \text{Poisson Clock}(\lambda_2)$$





Initial Phase

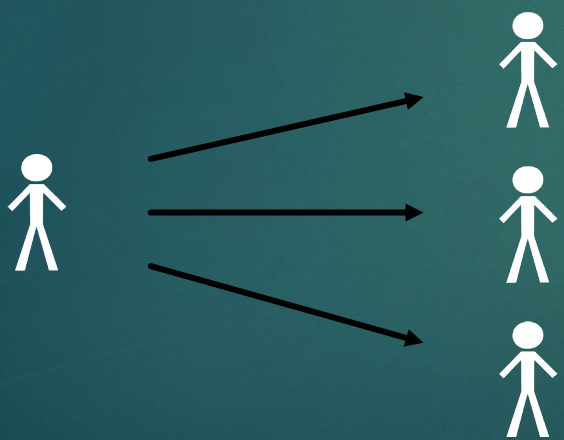


Kill

?

Share

Viral Phase



Inspect

?

Share

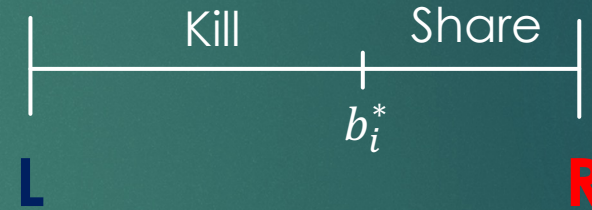
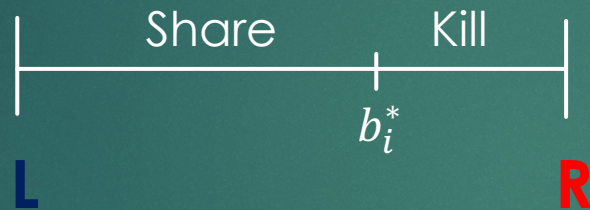
Equilibrium: Cutoffs



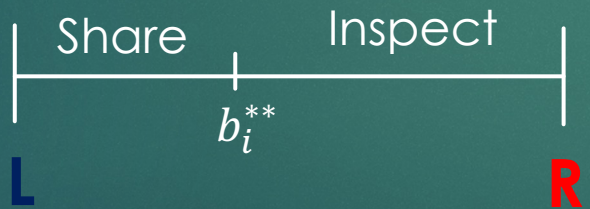
Left Article

Right Article

Initial Phase

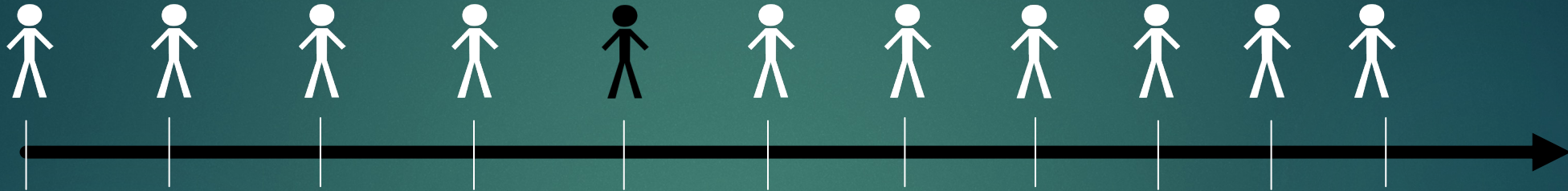


Viral Phase



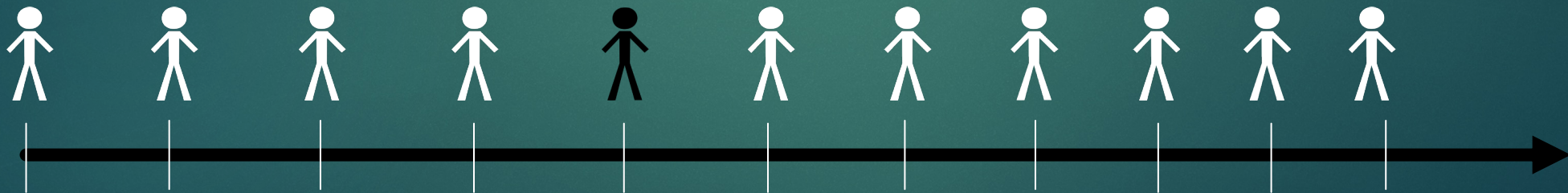
Equilibrium: Strategic Forces

Strategic Substitutes



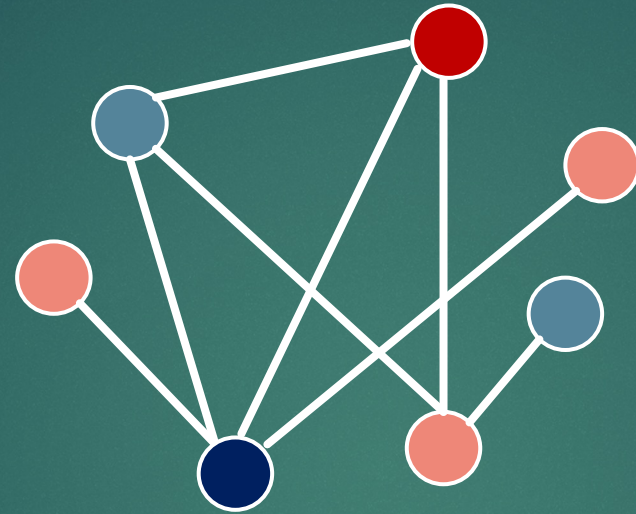
Inspect Inspect Share Inspect

Strategic Complements

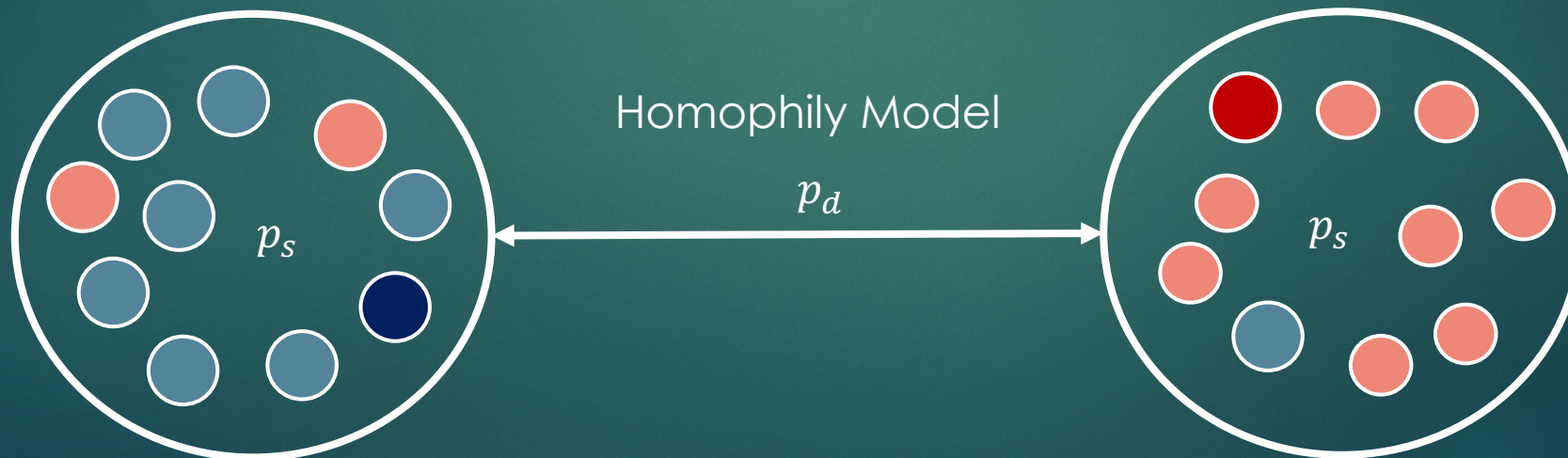


Inspect Share Inspect Inspect

Homophily is Bad for Misinformation



Uniform Connections

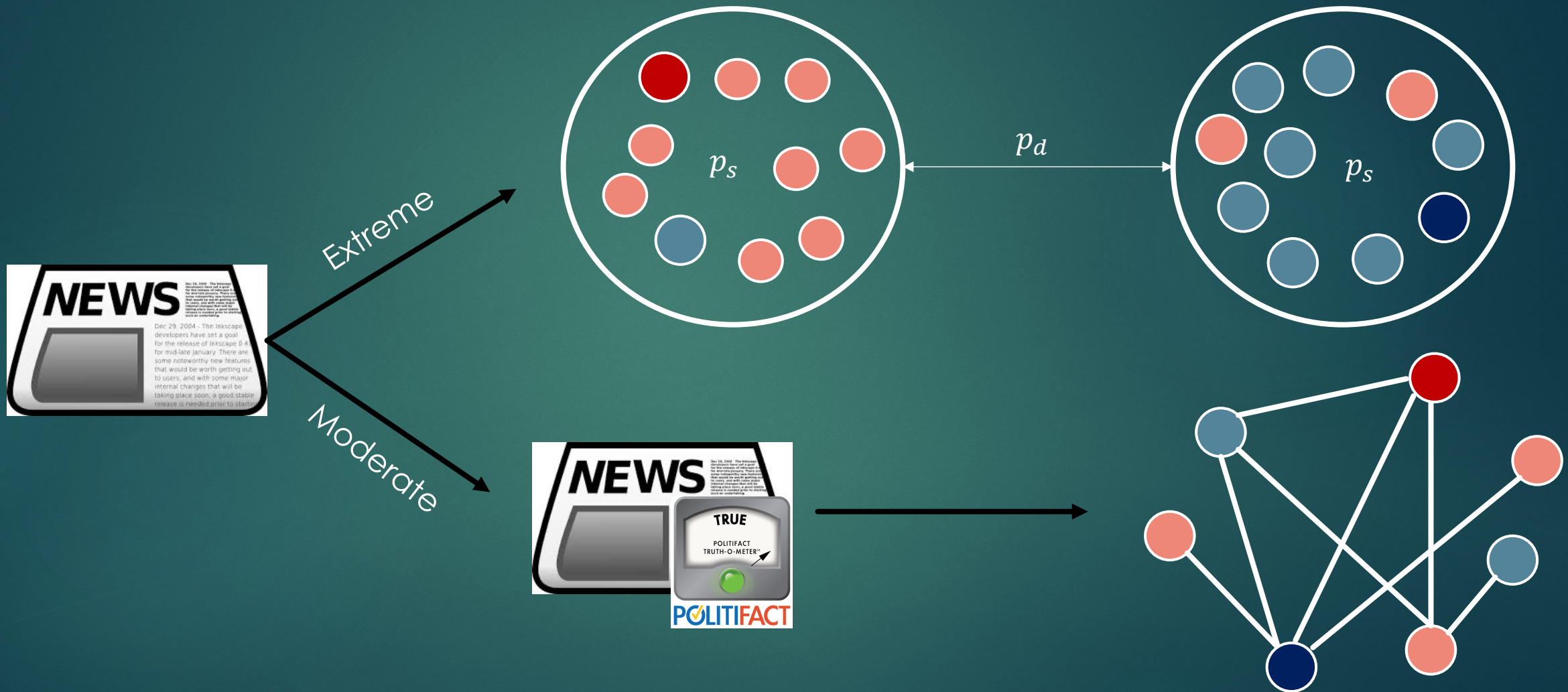


Platform Problem

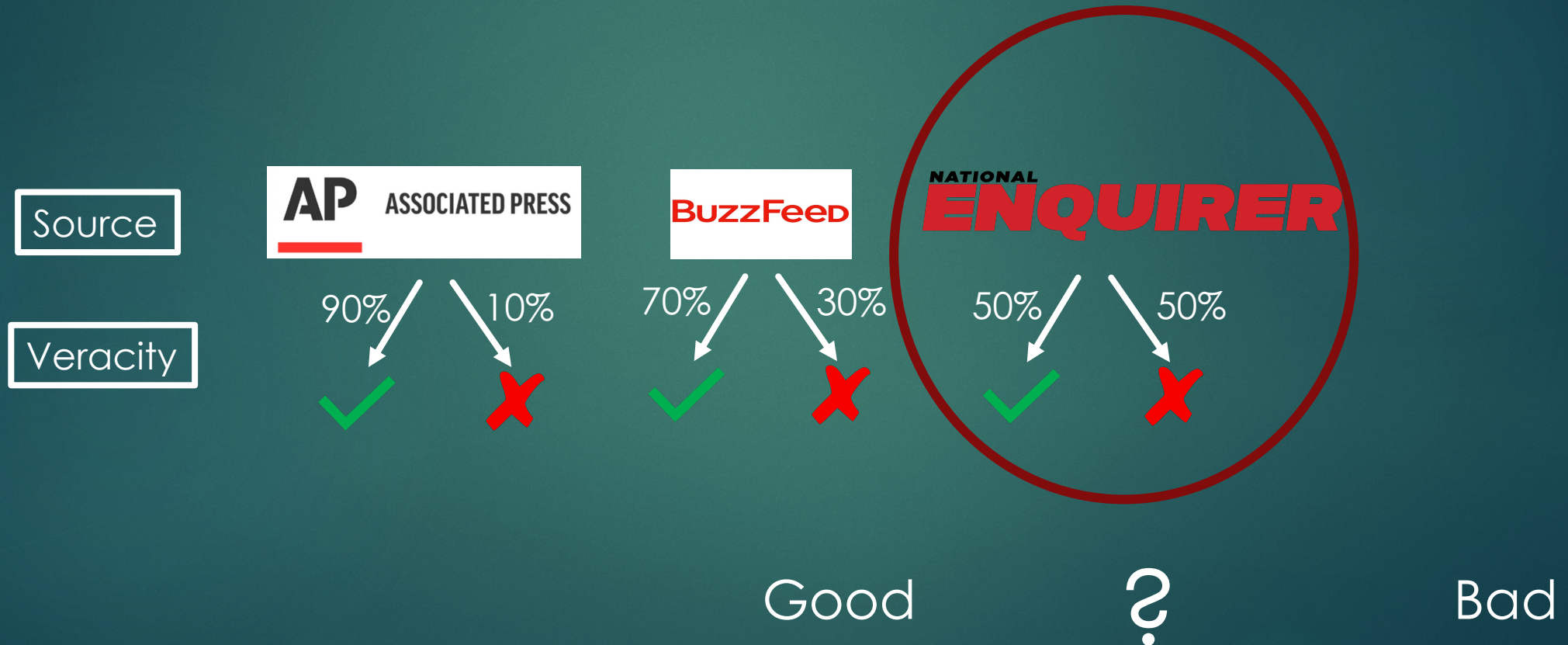


Recommendation

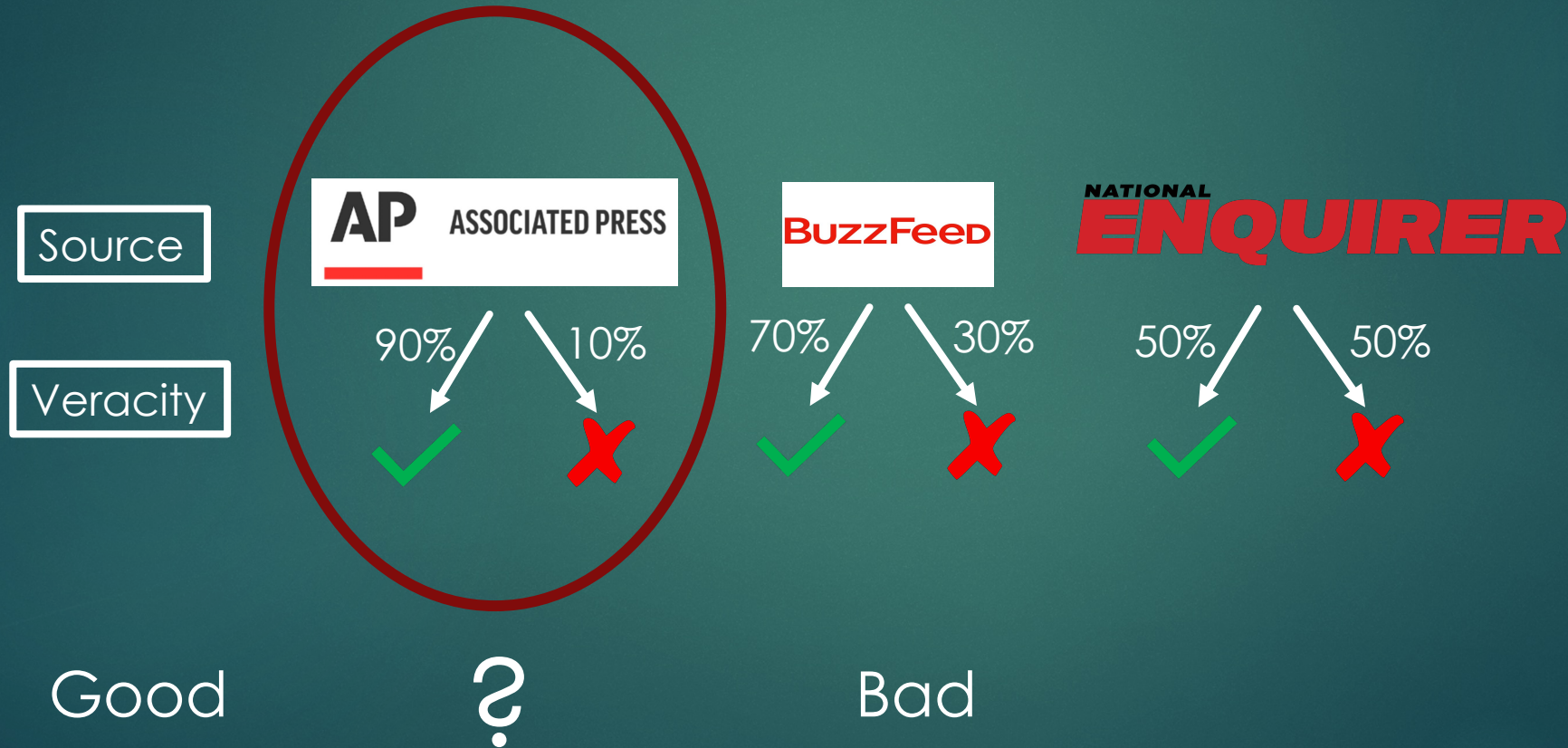
Filter Bubble Algorithm is Optimal



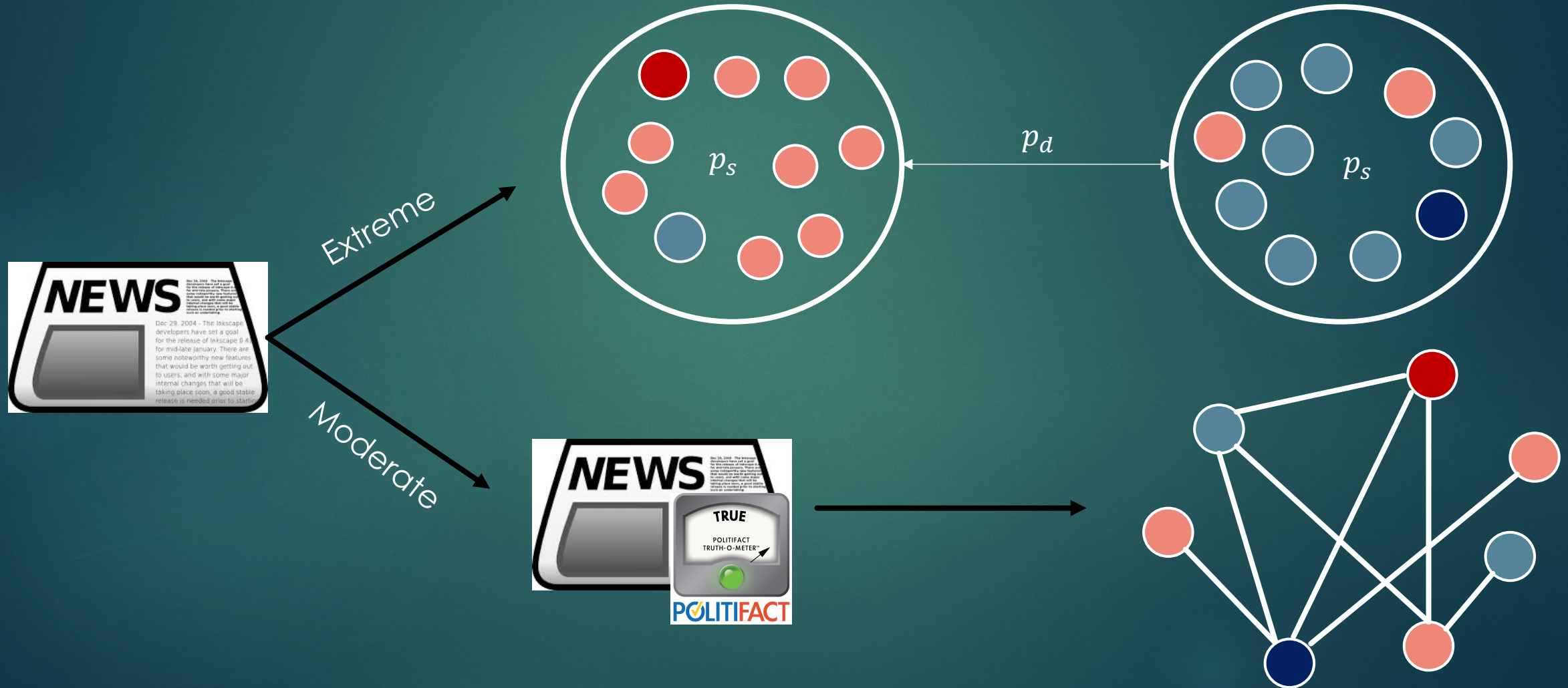
Combating Misinformation Spread: Provenance



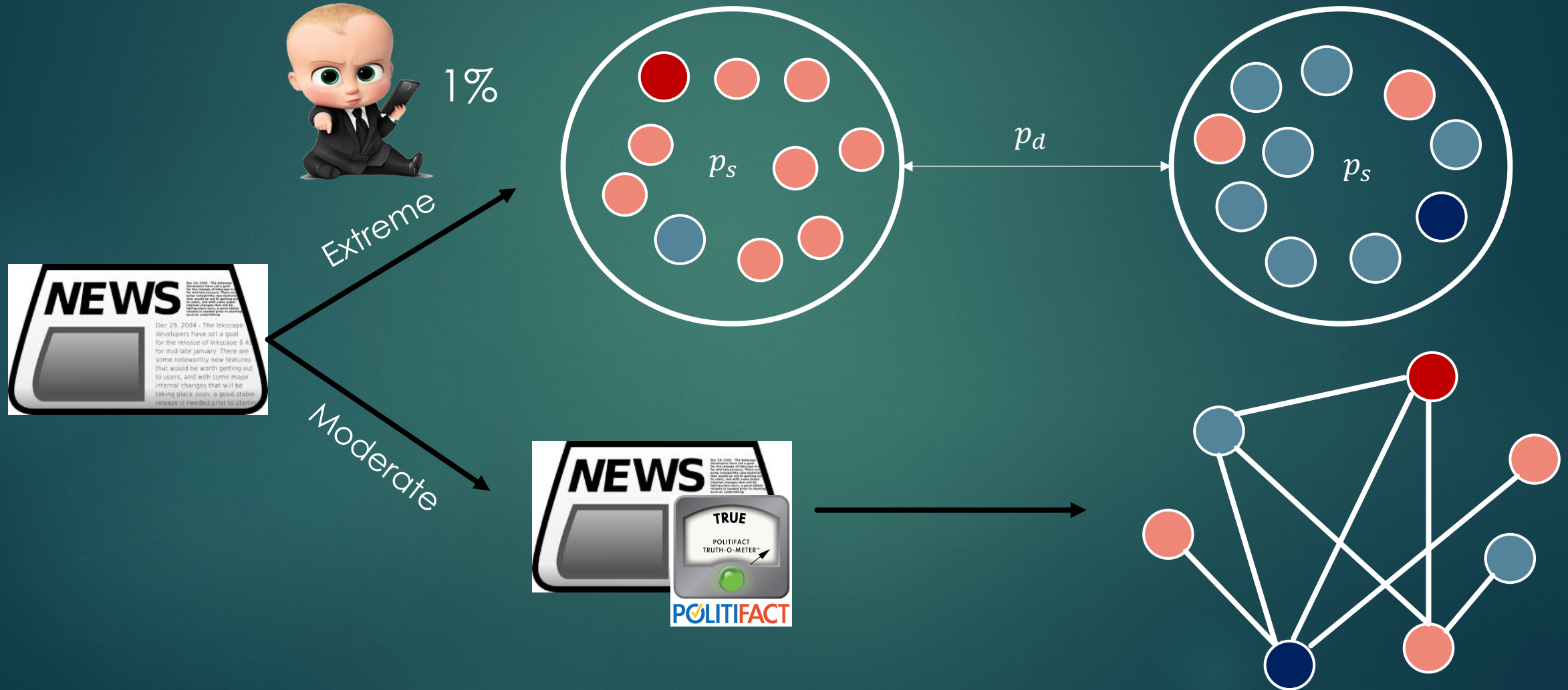
Combating Misinformation Spread: Provenance



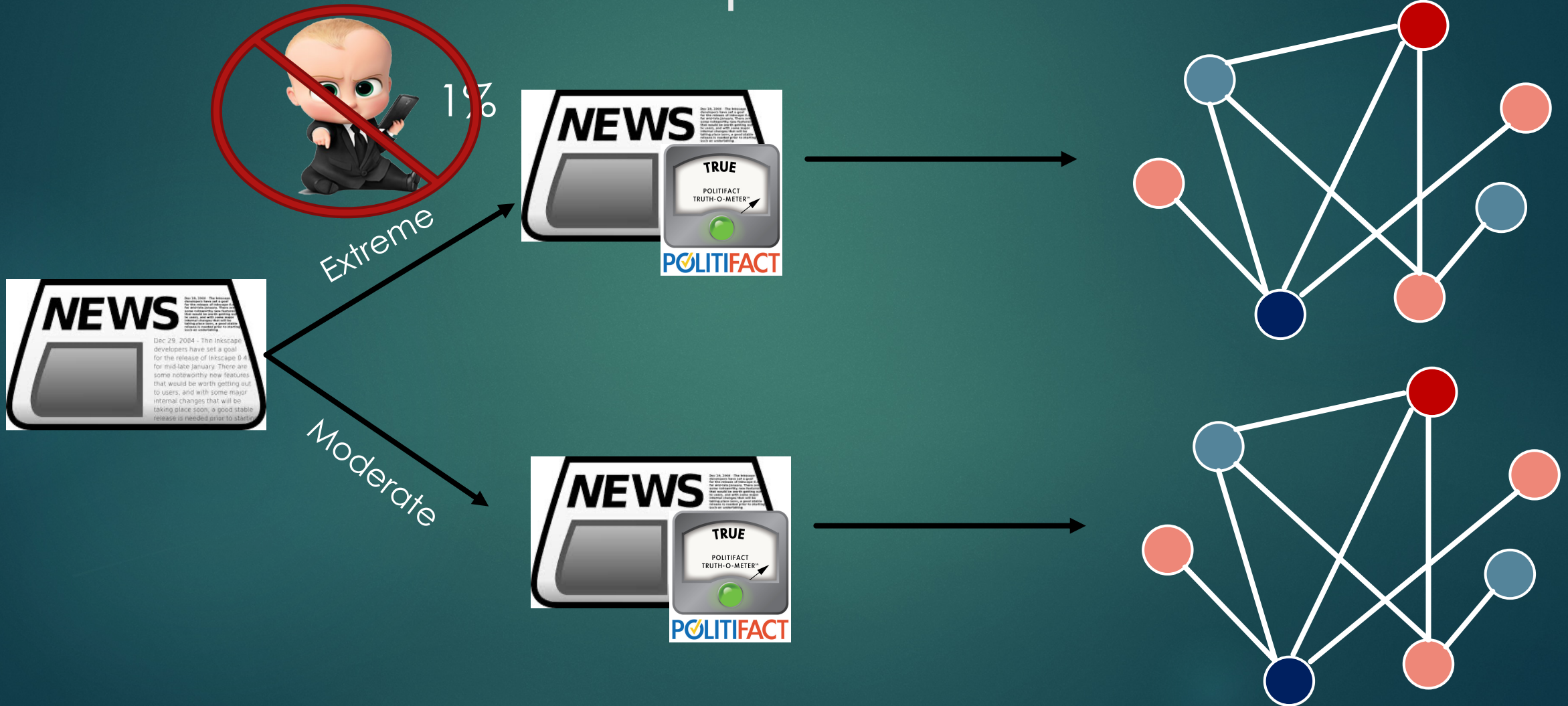
Combating Misinformation Spread: Threat of Censorship



Combating Misinformation Spread: Threat of Censorship

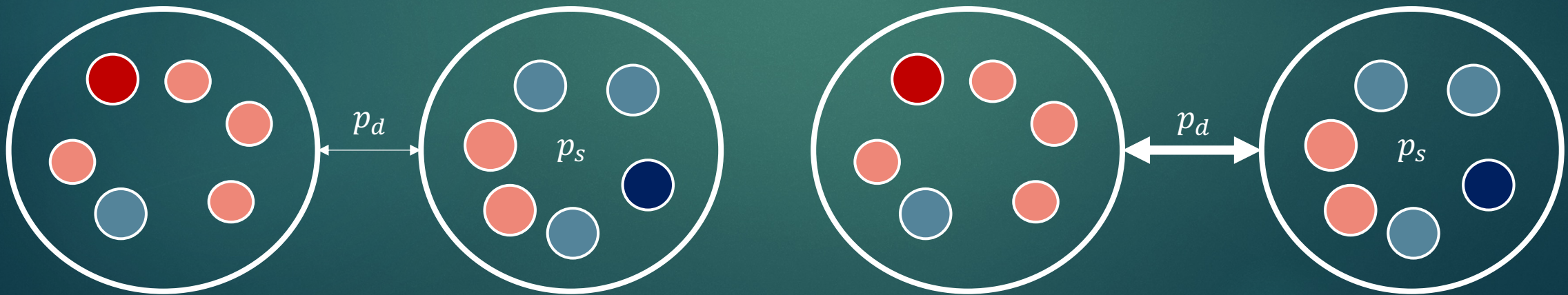


Combating Misinformation Spread: Threat of Censorship



Combating Misinformation Spread: Platform Algorithms

- ▶ Require that $\frac{p_s}{p_d} < \bar{p}$ for some \bar{p} that regulates the recommendation algorithm the platform can adopt.
- ▶ Highly-monotone, so $\bar{p} = 1$ not necessarily the optimal regulation, but $\bar{p} < \infty$ is.



Conclusion

- ▶ **Main tension:** the setting where content goes unchecked is exactly the setting where platforms should fact-check, but instead recommend unverified content.
- ▶ Do social media sites *have to* compromise engagement (e.g., ad revenue) to be “socially responsible”?
- ▶ Can we design “efficient” algorithms that allow users to have more agency over their content but do not propagate misinformation?

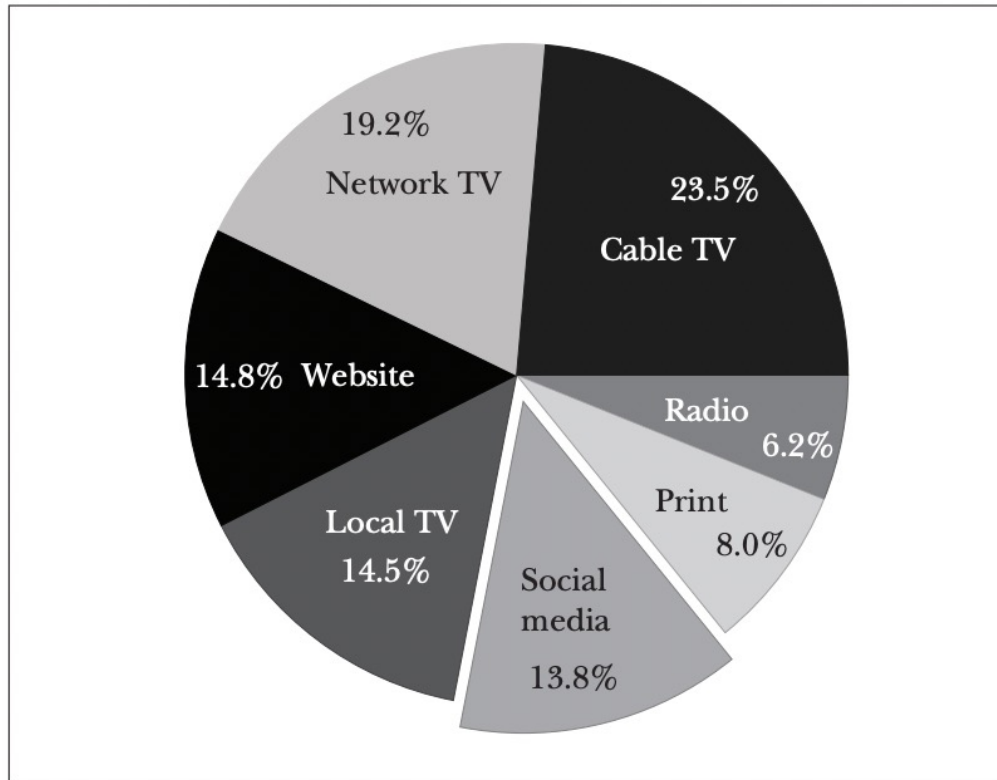


Misinformation: Strategic Sharing, Homophily, and Endogenous Echo Chambers

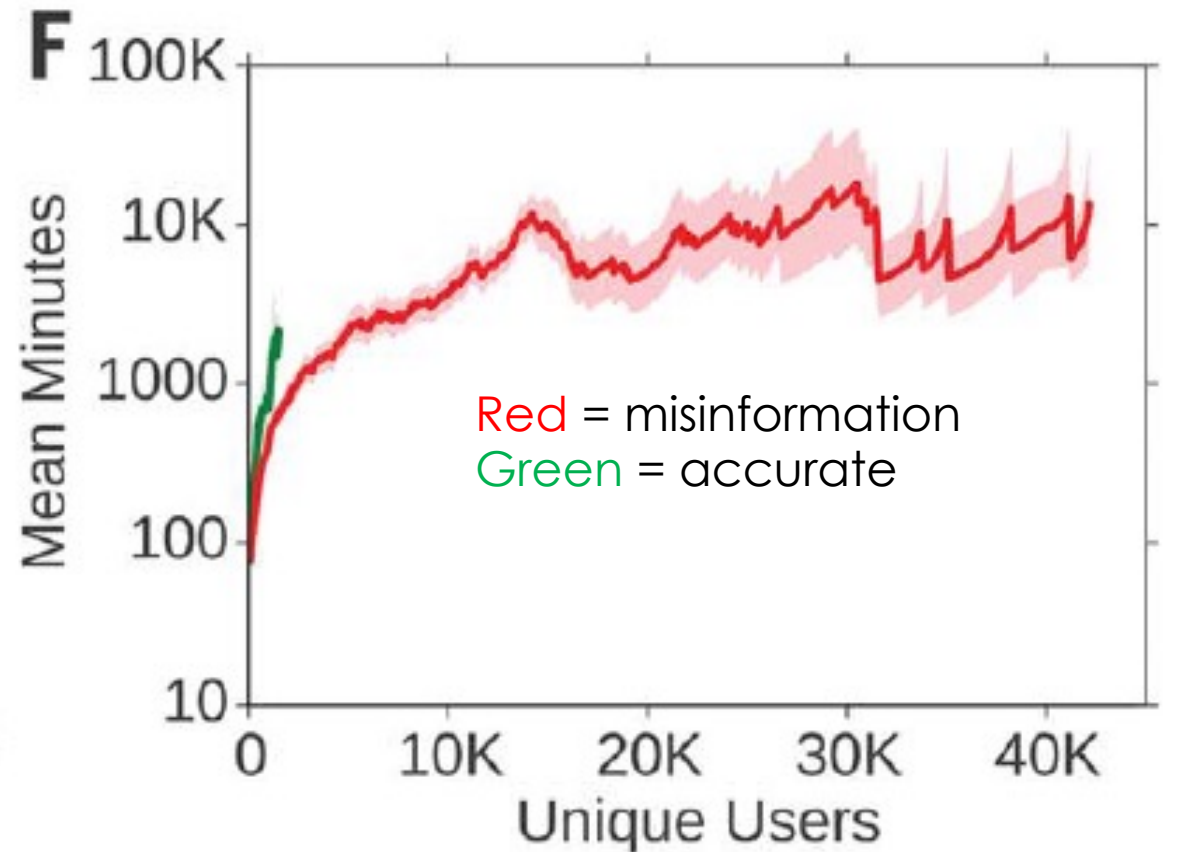
MODEL OF SHARING

Motivation: Platform Sharing

Most Important Source of 2016 Election News

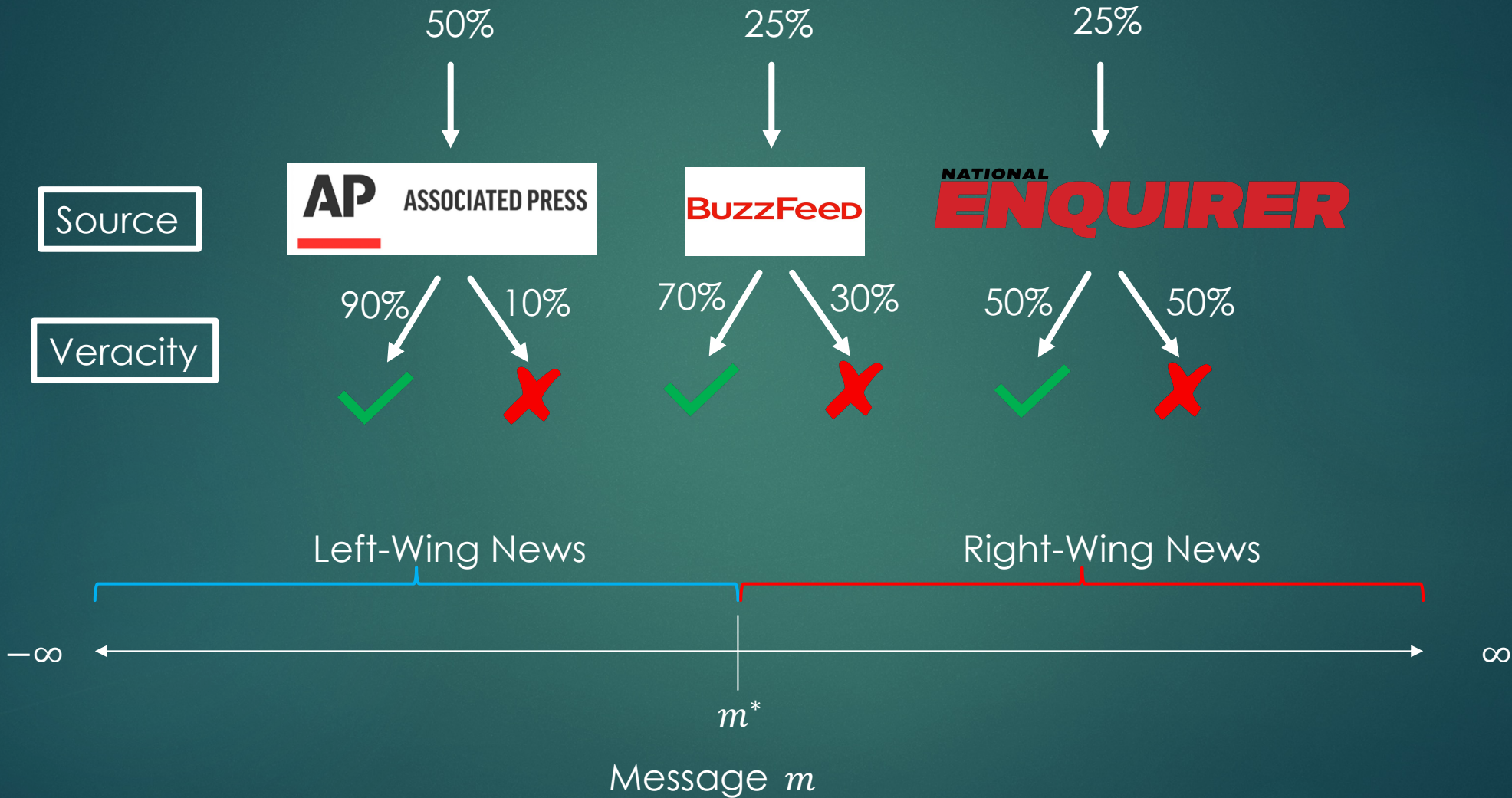


Allcott and Gentzkow (2017)

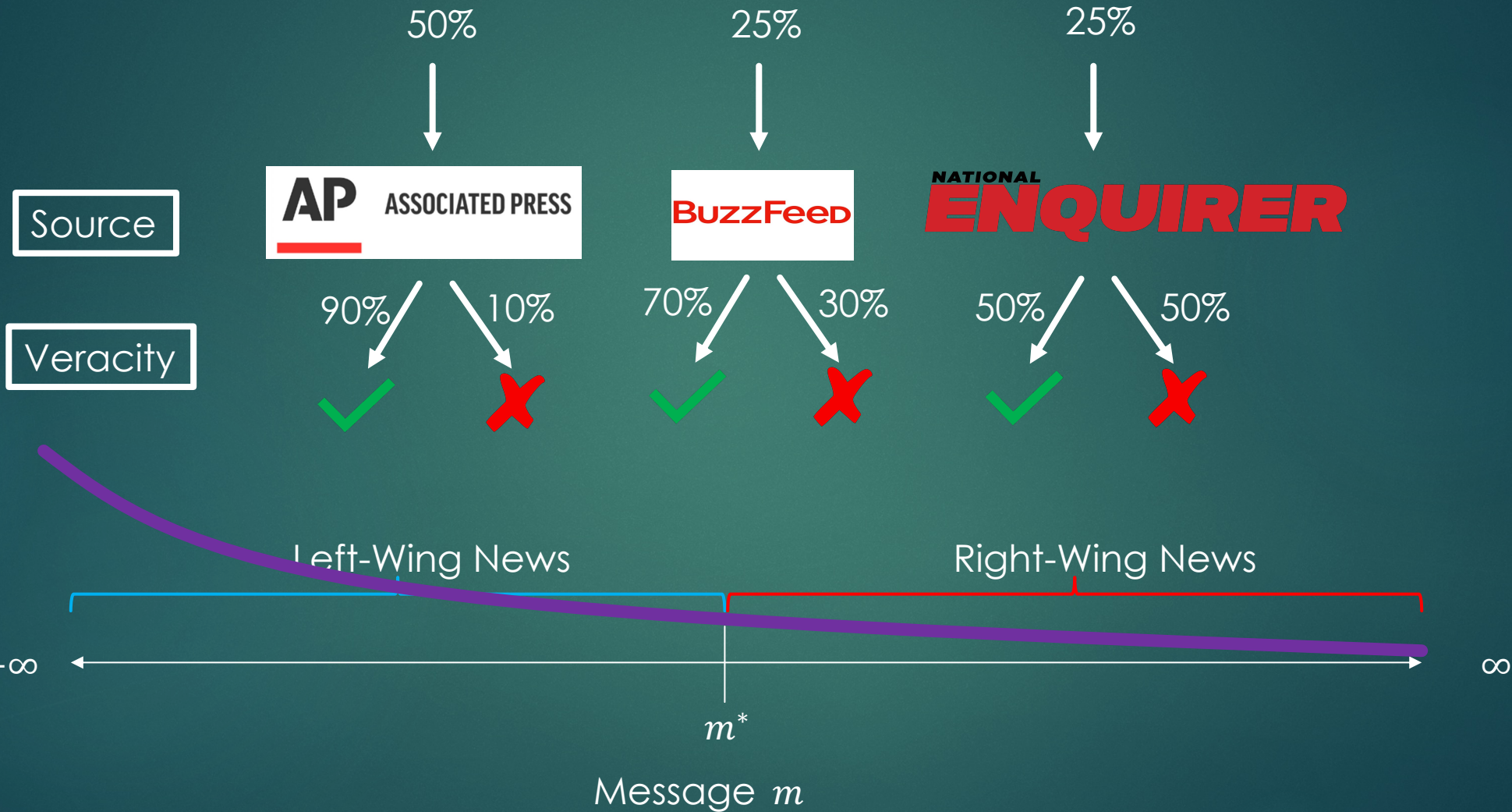


Vosoughi et al (2018)

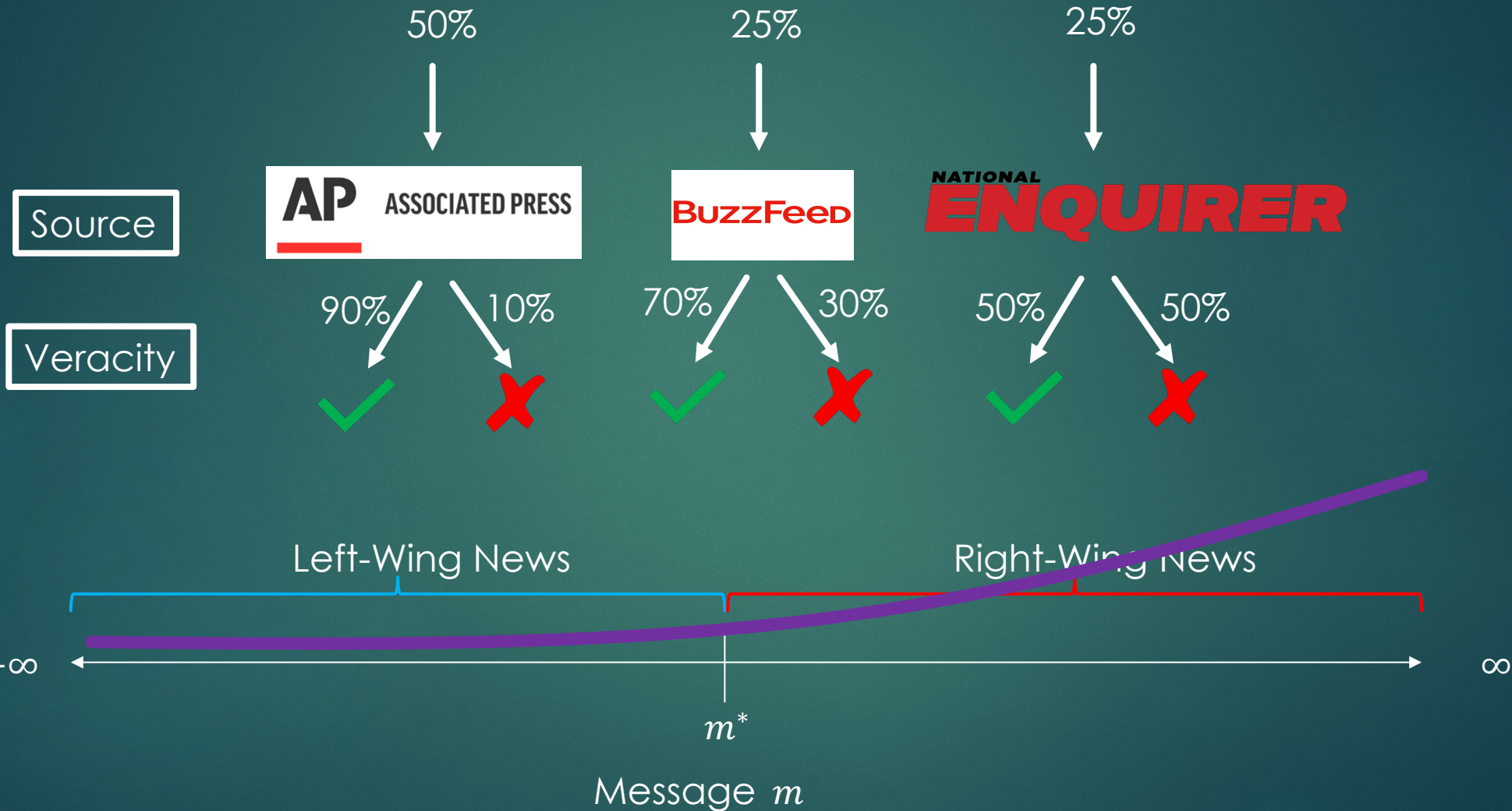
Model: News Generation



Model: News Generation



Model: News Generation



Model: Agents' Actions

Share



Inspect



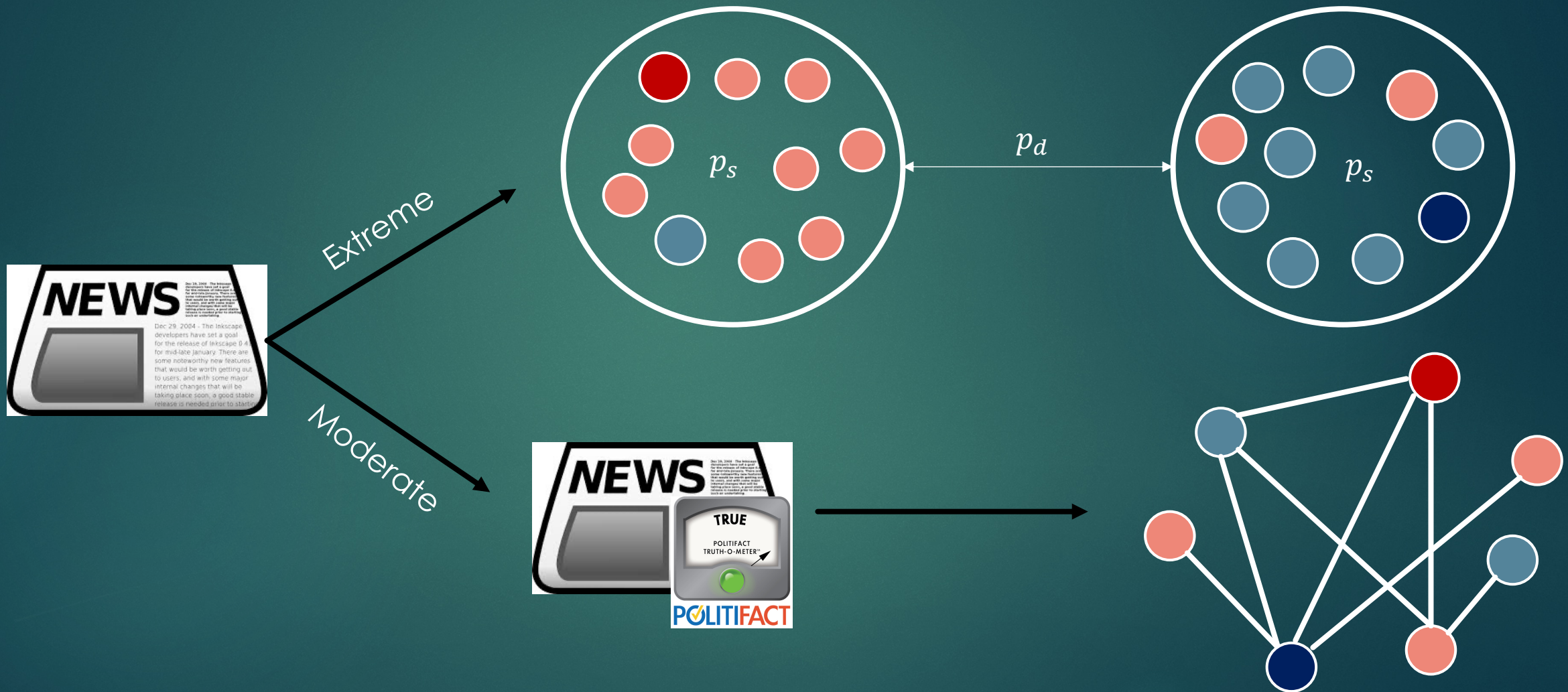
Kill



Platform Problem



Filter Bubble Algorithm is Optimal



Conclusion: Sharing Model

- ▶ The platform should choose the **sharing network** (through recommendations) to be one of two possibilities:
 - ▶ **(1) Extremist echo chambers with unverified content**
 - ▶ **(2) Diverse content with only verified content**
- ▶ How do we regulate platforms to push toward (2)?
 - ▶ **Provenance:** Show original sources of content
 - ▶ **Censorship:** Threaten to censor extreme unverified content
 - ▶ **Segregation Standard:** Require platform algorithms to spread cross-cutting content across ideologies