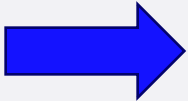


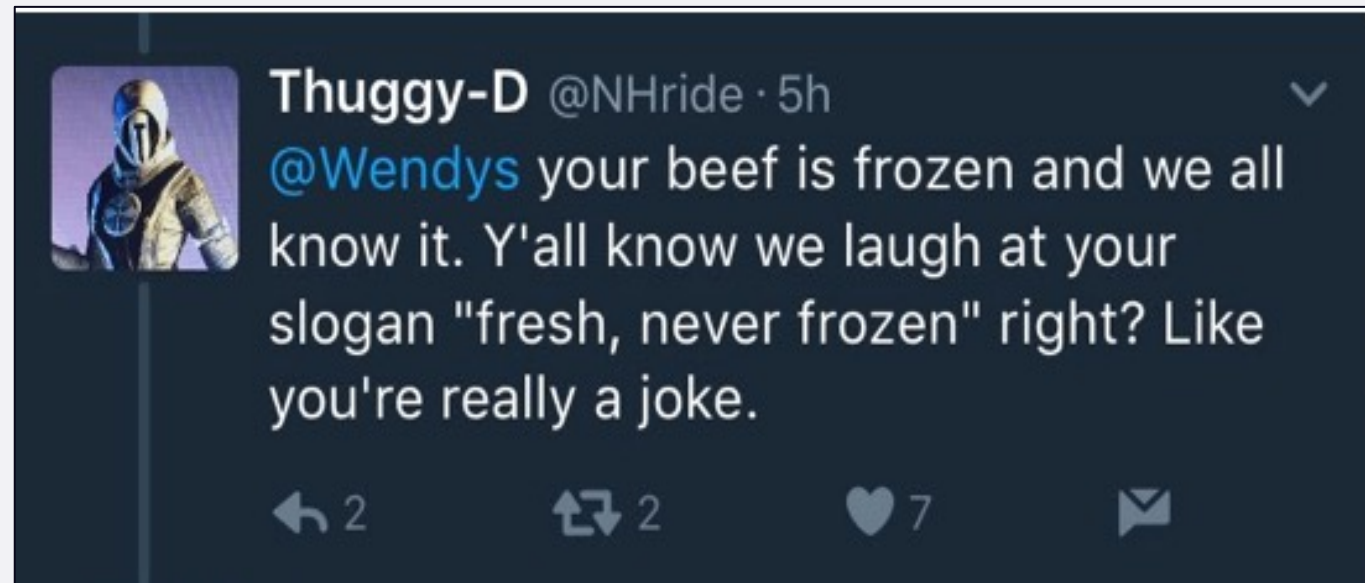
Research Challenges to Humanizing Cyberspace

Sara Kiesler, Carnegie Mellon University

25 July, 2023

Cyberspace - the good and the bad

- Trolls, from the beginning
- Deceptive advertising
- False complaints 
- Scams & ransomware
- Hate crimes
- Lies by celebs, politicians
- Conspiracy theories



Rumors and conspiracy theories are not new

The great fire of Rome

- Rumor - “Nero purposely caused the fire”
- Nero’s rumor - “Christians responsible”
- Many Christians killed



Today, conspiracy theories spread quickly online – MUCH faster than truth

“Pizzagate”

When: 2016 U.S. presidential campaign

What: Claim Podesta’s hacked email contains coded messages of human trafficking and child sex ring

How: Spread via 4chan, 8chan, Reddit, & Twitter



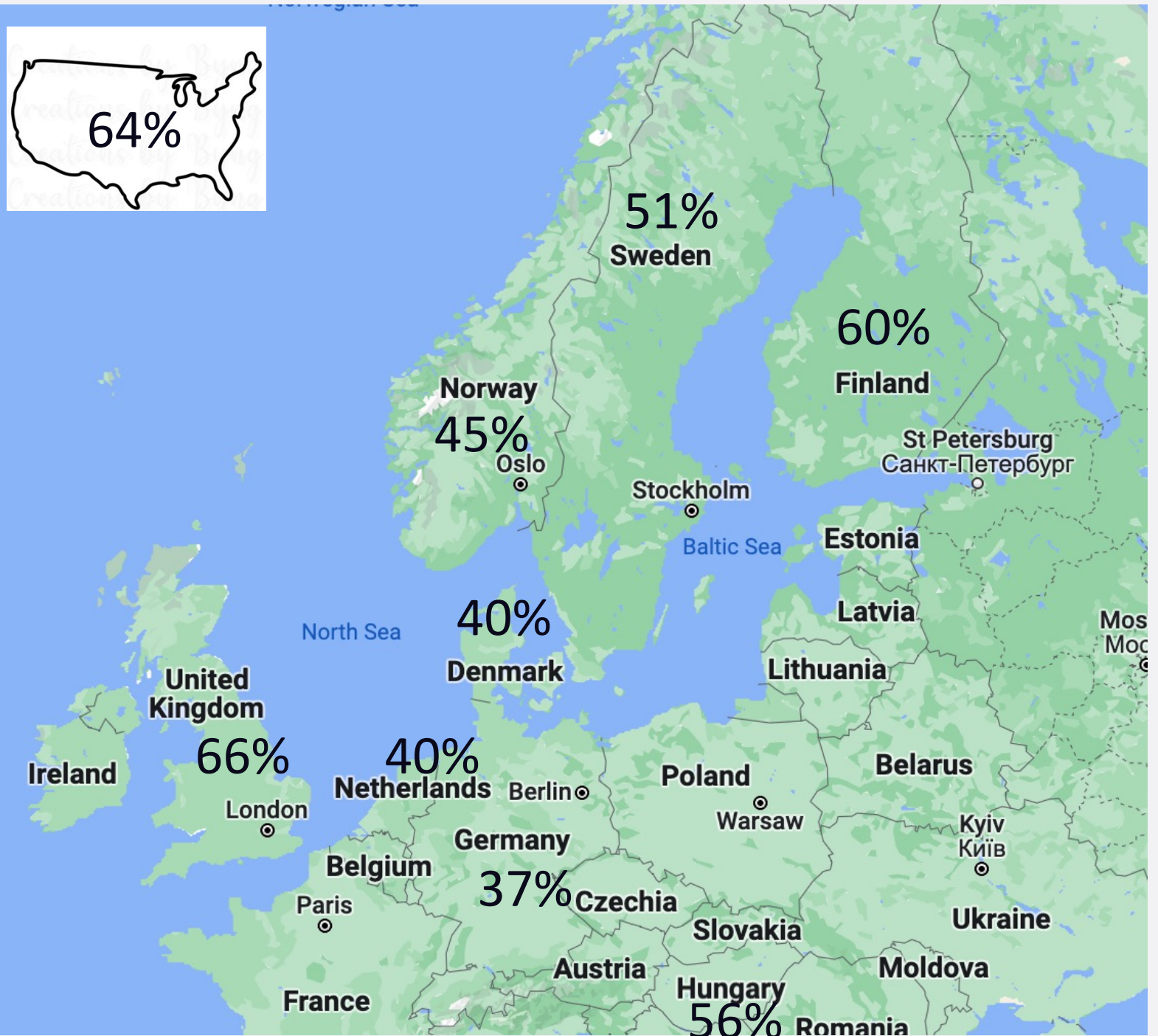
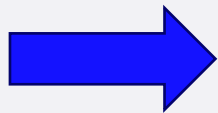
Comet Ping Pong Pizzeria



Worried Citizens

“Thinking about online news, I am concerned about what is real and what is fake on the internet.”

Agree or partly agree,
2021 survey



Disinformation

- Intentional falsehoods
- Disguised sources
- Spread assisted by AI architectures and algorithms

Retweets using Starbucks logo claims the coffee chain is hosting a "Dreamer Day" and offering free drinks to undocumented immigrants.

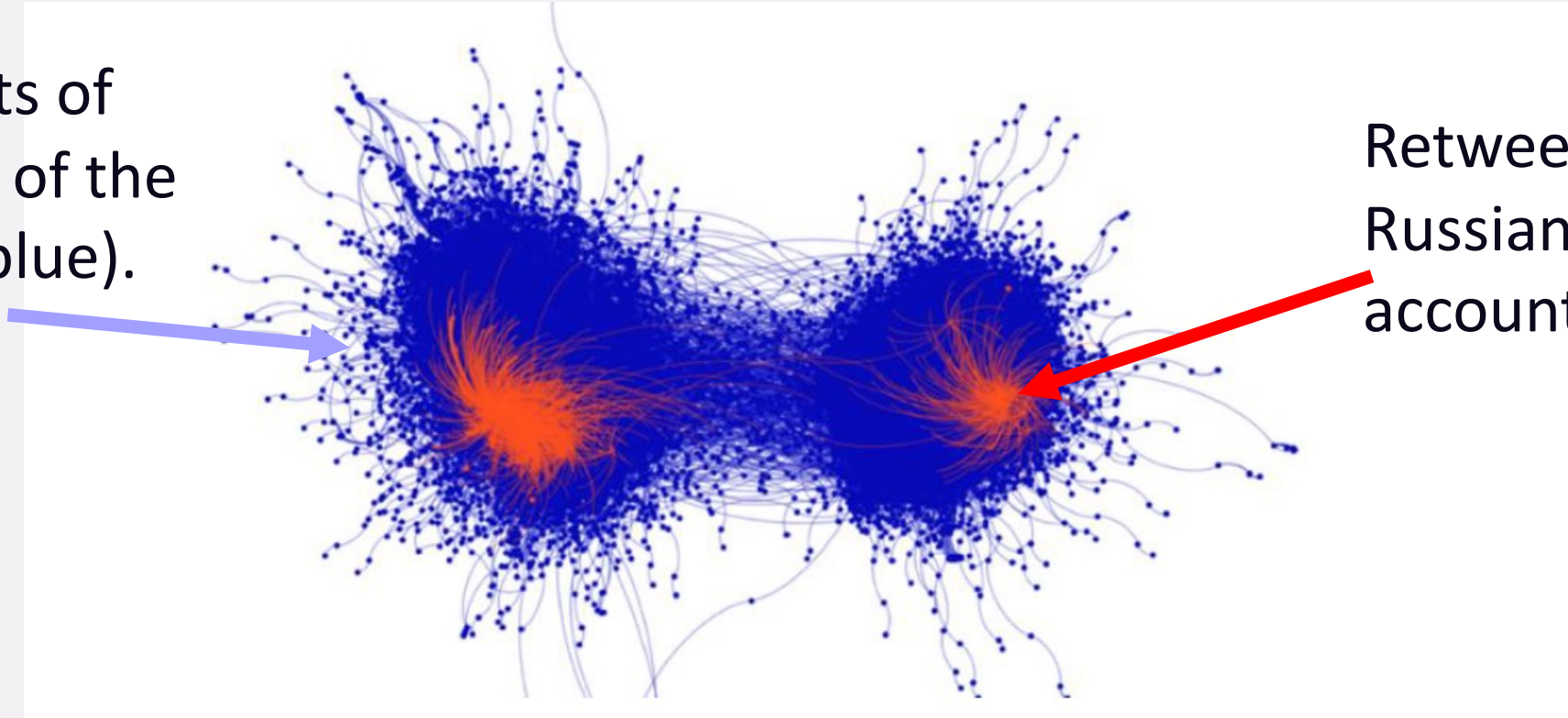
Starbucks is forced to reply:

Replying to [@01ProudAmerican](#)

This is completely false. Starbucks is not sponsoring any such event. Where did you get this information?

Disinformation campaign – Black Lives Matter

Retweets of the rest of the graph (blue).



Retweets of known Russian-IRA accounts (orange)

Why do people believe and share even patently false claims online? Such as..

SLAP (Secret Large-scale Atmospheric Program)

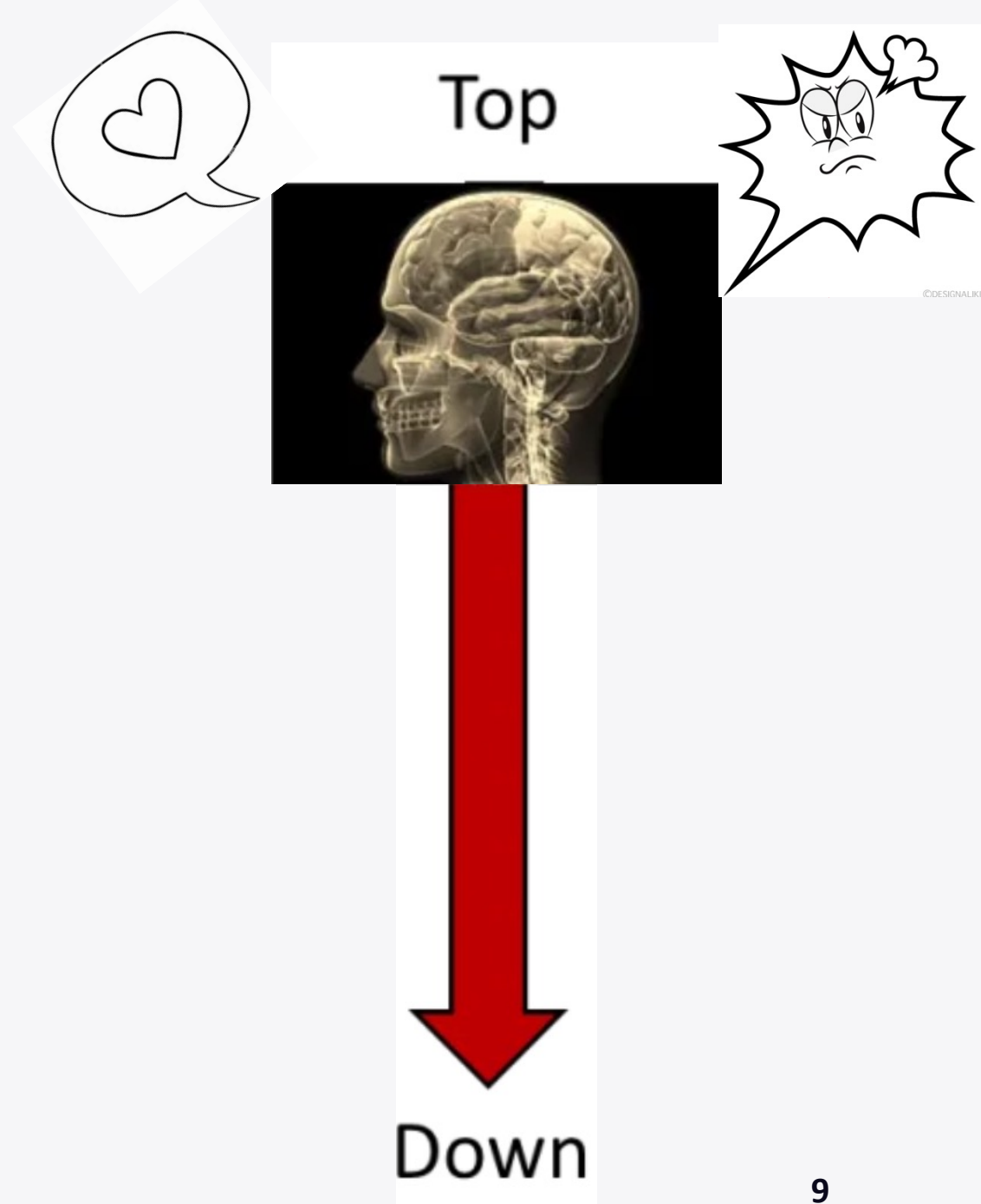
Condensation from aircraft
consist of toxic agents
Secret government policies!



Our top-down biases

- Limited, selective attention &
- Expectations and beliefs

Create bias



Choose



75% nonfat



25% fat

Choose



30% protein



70% carbohydrates & sugars

Top-down bias online

- Creates tunnel vision
 - Search leads people to believe what they searched for.
- Anchors new information
 - Most people think they can spot fake news. They can't.
- Creates overconfidence
 - Most people think others are biased but they are not.
 - People who say they understand computer concepts (even fake ones like “spindle transformations”) are vulnerable to phishing

Bottom-up biases

Please do not look at the blue letters!

Do not look at the blue letters

Raise your hand if you think this might be an accurate headline...

Brooms can stand up by themselves when the earth's gravitational pull is just right



Raise your hand if you think this might be an accurate headline...

Brooms can stand up by themselves when the bristles are spread just right



True?

Many of the recent wildfires in California, U.S. are caused by bad forest management.



True?

Many wildfires in California, U. S. are caused by downed electric power lines.



A test

I'm going to ask half of you to close your eyes while I show the first half something. **DON'T SAY WHAT IT IS!** (It's a secret from the folks with their eyes closed.)

Then I'm going to ask the first half to close their eyes and the second half to open their eyes, and I'll show them something. **DON'T SAY WHAT IT IS!** (It's a secret from the folks with their eyes closed.)



Those with eyes open, now close.
Those with eyes closed, now open.



OK, New Rules

- Everybody open your eyes
- Now I'll show you all something and I want you to shout out what it is



What's going on?

- What we believe is influenced by our past experience with the world
- The line between memory and perception is blurry!



What Group 1 saw

What all saw at test

What Group 2 saw

Social biases



“I’ve never trusted cows.”

We distrust others
unlike ourselves



We seek people like
ourselves



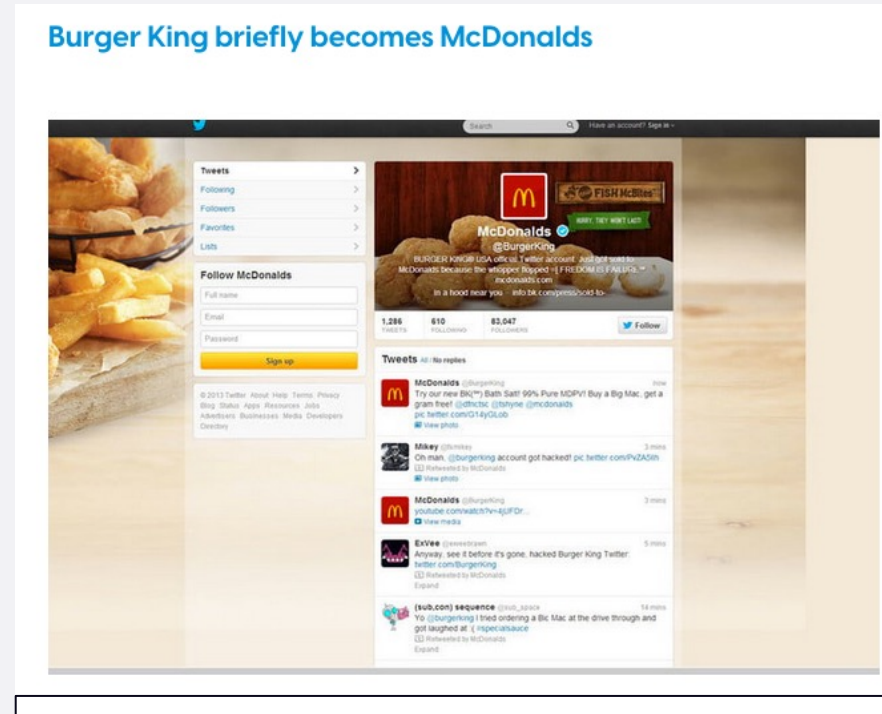
Platform algorithms



Polarization online

Why do people *spread* misinformation?

- Personal experience
 - Dissatisfactions
 - Slight
- Interpersonal
 - Group identity
 - Fun, novelty
 - Reputation
 - Profit



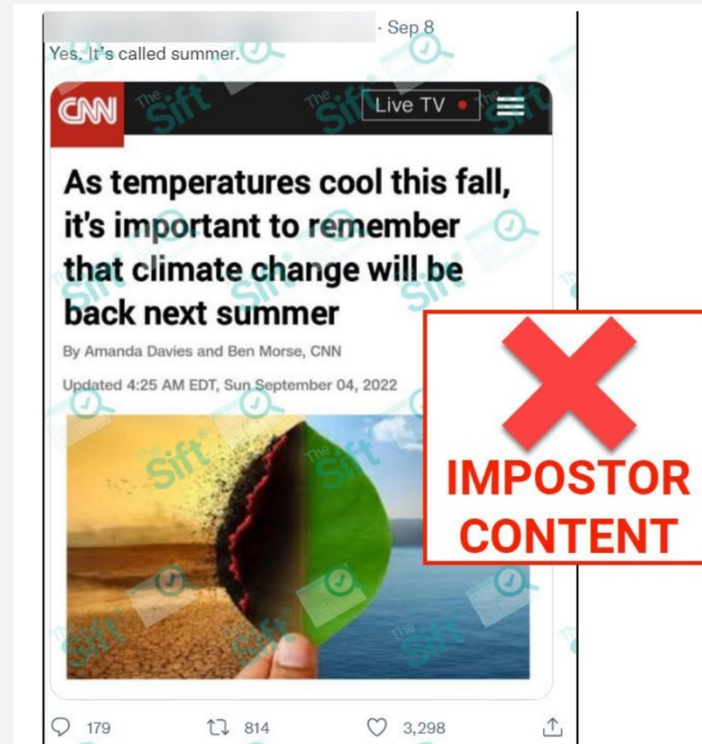
The hack of Burger King's Twitter feed plastered unpleasant tweets all over the Burger King feed, & announced Burger King was sold to McDonald's "because the whopper flopped," and even changed the logo to that of McDonald's.

Collective motivations

- Moral convictions
 - Anti-vaccine
 - Anti-abortion
 - Climate denial
 - Defense of religion
 - Defense of country
 - Defense of partisanship



Tweet w. 949 likes in less than a day



TikTok post of false CNN headline

Corporate & state motivations

- Corporate sponsored
 - Disparage competitors
 - Stave off regulation
- State sponsored - global
 - Create uncertainty, chaos around elections and foreign policy

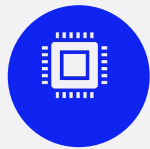


“We want the demilitarization and denazification of Ukraine.”

If misinformation is fundamental to human nature, what can be done?



Improving
technology



Better data
from
platforms



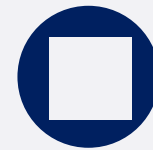
Improving
debunking
and
inoculations



Repairing
scientific
publishing



Legal or
market
solutions



Educating
the public

Improving technology: Advances in detection and labeling

Better detection → earlier labelling

Target Deepfake



Caricature



Example: Debunking by distorting the deepfake

Detecting conspiracy theories, modeling code words

Aliens

Big-pharma

Bio-engineering

Chemtrails

Corporate media

Deep State

Fake science

Flat earth

False flag

George Soros

Globalist

GMO

Illuminati

Media

New World order Cabal

Nutritional supplements

Pedophile rings

Rothchilds

Vax

Zionist

9-11 truther



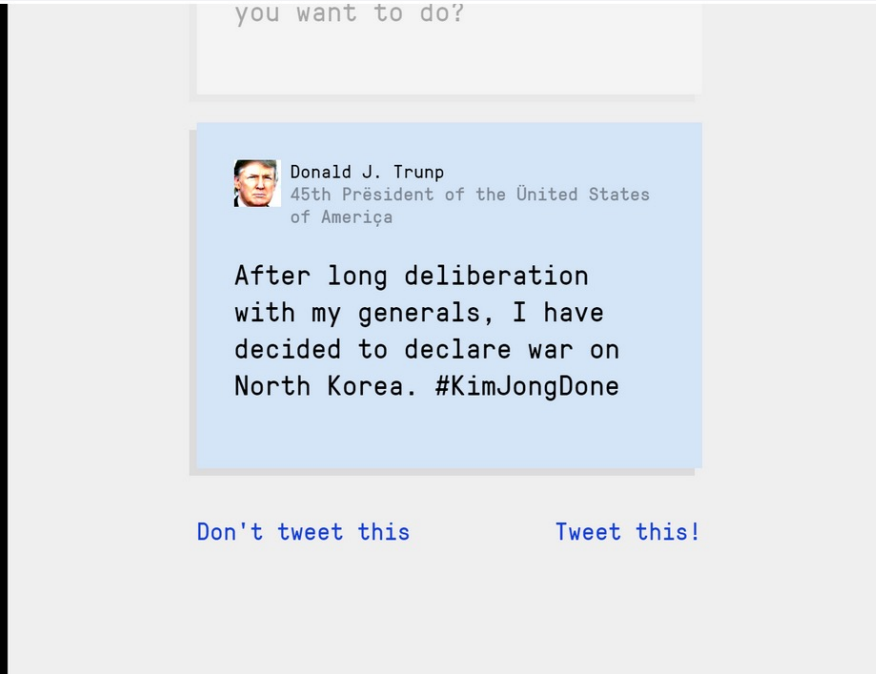
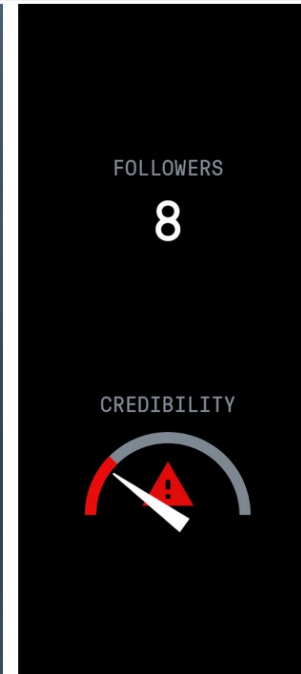
Debunking

“Many of the recent wildfires in California, U.S. are caused by bad forest management.”



Checked by independent fact checkers

Inoculation



Landing page of bad news game

Improving the scientific information ecoystem

- Scientific rewards based on timeliness and excitement
 - In China, prizes and bonuses
 - In U.S., tenure and grants
- Predatory publishing
 - Authors pay
 - Encourages shaky results
- News articles do not cite sources or questions
 - Media attention given to “breakthroughs”
 - News headlines appeal to emotion and hope



Solving scientific error and misinformation

- Change the system of rewards
 - fewer papers
 - higher quality
- Identify errors, misuse of statistics, and biases in analysis
- Do not reward scientists for publishing in predatory journals
- Educate and pay more science writers

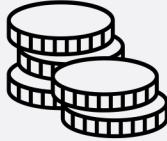


Legal solutions?

- Bans on content may threaten free speech, art, education, and state security.
- Civil liability difficult and costly
- Sue the platform?



Market solutions

- Private sector likely to sell services to prevent harms—like the identity protection, location verification, and virus prevention services of today.
 - Immutable life logs or authentication trails. 
 - Proofs of your authorship 
 - Incentivized certification programs 
- Platform speech accountability policies reform and misinformation tracking

Conclusion

- The bad and harmful in cyberspace reflects human nature
- Technological fixes will help
- Legal and market solutions will help those who care
 - How do we get people to care?
- Dire need for education of all ages and cognitive tools to aid critical thinking



Factshala in India

~~Research Challenges to Humanizing Cyberspace~~

Research Challenges towards Civilizing Cyberspace

The End

Sara Kiesler

kiesler@cs.cmu.edu

References

Slide 3. https://en.wikipedia.org/wiki/Great_Fire_of_Rome;
[https://commons.wikimedia.org/wiki/File:Robert, Hubert - Incendie %C3%A0 Rome - .jpg#/media/File:Robert, Hubert - Incendie à Rome -.jpg](https://commons.wikimedia.org/wiki/File:Robert,_Hubert_-_Incendie_%C3%A0_Rome_-_jpg#/media/File:Robert,_Hubert_-_Incendie_%C3%A0_Rome_-_jpg)

Slide 5. B. Johansson, Ø. Ihlen, J. Lindholm, & M. Blach-Ørsten. Eds *Communicating a pandemic: Crisis management and Covid-19 in the Nordic countries.*. Nordicom, University of Gothenburg , 2023, p. 303-324 <https://doi.org/10.48335/9789188855688-14>

Slide 6. For definitions and more information see: <https://www.whitehouse.gov/wp-content/uploads/2022/12/Roadmap-Information-Integrity-RD-2022.pdf>

Slide 7. Ahmer Arif, Leo G. Stewart, and Kate Starbird. 2018. Acting the Part: Examining Information Operations Within #BlackLivesMatter Discourse. In *Proceedings of the ACM on Human-Computer Interaction*, Vol. 2, CSCW, Article 20 (November 2018). ACM, New York, NY. 26 pages. <https://doi.org/10.1145/3274289>(November 2018). ACM, New York, NY. <https://doi.org/10.1145/3274289>. See also Kate Starbird's website for more references.

Slide 9, 13. <https://owlcation.com/social-sciences/Perception-in-Psychology>

Slides. 14-17.and 32. Courtesy of Chris Wahlheim. See also Wahlheim, C. N., Alexander, T. R., & Peske, C. D. (2020). Reminders of everyday misinformation statements can enhance memory for and beliefs in corrections of those statements in the short term. *Psychological Science*, 31,10,1325.

References continued.

Slides 18-24. Bugelski B.R., Alampay D.A.. (1961). The role of frequency in developing perceptual sets. *Canad. J. Psychol.*, 15, 205-211.

Slide 25. Cartoon by permission from The New Yorker Collection 1997. Warren Miller from cartoonbank.com.

Slide 26. Soroush Vosoughi et al., The Spread of True and False News Online, 359 *Science* 1146, 1146 (2018), <http://science.sciencemag.org/content/359/6380/1146/tab-pdf> [<https://perma.cc/5U5D-UHPZ>]. Also slide 4.

Slide 30. Fosco, C., Josephs, E., Andonian, A., Lee, A., Wang, X., & Oliva, A. (2022). Deepfake caricatures: Amplifying attention to artifacts increases deepfake detection by humans and machines. *arXiv preprint arXiv:2206.00535*.

Pu Sun, Yuezun Li, Honggang Qi and Siwei Lyu. Faketracer: Exposing Deepfakes With Training Data Contamination. In IEEE Conference on Image Processing (ICIP), Bordeaux, France, 2022

Slides 32-33. Stephan Lewandowsky & Sander van der Linden (2021) Countering Misinformation and Fake News Through Inoculation and Prebunking, *European Review of Social Psychology*, 32:2, 348-384, DOI: [10.1080/10463283.2021.1876983](https://doi.org/10.1080/10463283.2021.1876983)

Ecker, U.K.H., Lewandowsky, S., Cook, J. *et al.* The psychological drivers of misinformation belief and its resistance to correction. *Nat Rev Psychol* 1, 13–29 (2022). <https://doi.org/10.1038/s44159-021-00006-y>

References continued

Slides 32-33 continued. Roozenbeek, J., van der Linden, S., Nygren, T. (2020). Prebunking interventions based on “inoculation” theory can reduce susceptibility to misinformation across cultures. *Harvard Kennedy School Misinformation Review*. <https://doi.org/10.37016//mr-2020-008>

Slides 34-35. Jevin D. West & Carl T Bergstrom.(2021). Misinformation in and about science.Proceedings of the National Academy of Sciences, 118, 15. PubMed: [33837146](https://pubmed.ncbi.nlm.nih.gov/33837146/). <https://orcid.org/0000-0002-4118-0322>
jevinw@uw.edu

Zuo, C., Acharya, N., & Banerjee, R. Querying across genres for medical claims in news. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, pages 1783-1789. Nov. 2020.

Y Zhao, X He, Z Feng, S Bost, M Prosperi, Y Wu, Y Guo... (2022). Biases in using social media data for public health surveillance: A scoping review.- International Journal of Medical Informatics

Slide 36-37. Chesney, Robert and Citron, Danielle Keats, Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security (July 14, 2018). 107 California Law Review 1753 (2019), U of Texas Law, Public Law Research Paper No. 692, U of Maryland Legal Studies Research Paper No. 2018-21, Available at SSRN: <https://ssrn.com/abstract=3213954> or <http://dx.doi.org/10.2139/ssrn.3213954>

Slide 38. Factshala <https://factshala.com/>