Platforms, Free Speech and the Problem of Fake News

Marshall W. Van Alstyne
July 2021

Draft 0.5 – This draft is circulated for comments and critique. Please do not quote or cite without checking first. Praise, flaws, critiques etc. are all welcome and invited at mva@bu.edu.

Abstract:

How should a platform or a society address the problem of fake news? The spread of misinformation is ancient, complex, yet increasingly present in Asian, European, Latin American, and US elections. After examining key attributes of "fake news" and of current solutions, this article presents tradeoffs in the design of a Fair News Distribution Mechanism. The challenges are not restricted to truth or to scale alone. Surprisingly, there exist boundary cases when a just society is better served by a mechanism that allows lies to pass, even as there are alternate boundary cases when a just society should put friction on truth. Harm reflects an interplay of lies, decision error, scale, and externalities. Single factor solutions fail on multifactor problems. Using mechanism design, this article then proposes three tiers of solutions: (1) those that are legal and business model compatible, so firms should adopt them (2) those that are legal but not business model compatible, so firms need compulsion to adopt them, and (3) those that require changes to bad law. The article concludes by proposing tests for the legitimacy of various interventions on free speech.

Introduction

Fake news is a problem. It is a near universal problem.[1] Trolls and propagandists have used it as a weapon in political campaigns, anti-vaccination campaigns, nutrition battles, insurrections, and to sow ethnic conflict. The inventor of the World Wide Web identified fake news as one of the three most dangerous assaults on the Internet.[2] Fake news is not new. Stone frescoes record the victories of "Rameses the Great" on temples from the 13th century BC yet more complete records of the treaty between Egyptians and Hittites show his battle was a stalemate.[3]

[1] Governments across all seven continents have been affected. The Wikipedia entry on fake news lists accounts of problems in Australia, Austria, Brazil, Canada, Czech Republic, China, Finland, France, Germany, India, Israel, Malaysia, Mexico, Myanmar, Netherlands, Pakistan, Palestine, Philippines, Poland, Singapore, South Africa, South Korea, Spain, Sudan, Sweden, Syria, Taiwan, Ukraine, United Kingdom, United States, and Venezuela. https://en.wikipedia.org/wiki/Fake_news Accessed August 18, 2018.

[2] His other two are citizen surveillance and cyber-warfare. The World Wide Web's inventor warns it's in peril on 28th anniversary By Jon Swartz, *USA Today*. March 11, 2017.

[3] Weir, William (2009). *History's Greatest Lies*. Beverly, Massachusetts: Fair Winds Press. pp. 28–41.

Fake news is a known problem. Figure 1 shows occurrences of the terms "false news," "fake news," "misinformation," and "disinformation" in literature since 1800. Until recently, misinformation has been the more common term. Unsurprisingly, "misinformation" spiked during both World War I and World War II as conflicting powers sought to demoralize each other's troops with concocted stories of false victories.[4] "Disinformation" spiked during the Cold War between the Soviet Union and the United States, then rose again under the administration of Vladimir Putin.[5] "Fake news" was common under Donald Trump. Deceit is common among adversaries and those pushing hidden agendas.

Challenges of modern fake news, however, render it even more pernicious than in the past. One reason is that modern platforms allow secret or insulated public messaging. The propagandist can whisper his case at a scale that is simultaneously vast yet almost invisible to those who, on observing the message, would oppose it with countervailing evidence. During the 2016 U.S. presidential election, members of the Trump campaign could buy ads targeted at individual coal miners in Pennsylvania.[6] The timing and content of such messages could remain hidden from the Clinton campaign in a manner that is largely impossible using traditional broadcast media.[7] A second reason modern platforms exacerbate the problem is inclusivity. Like most technology, social technology cuts two ways. Modern platforms give voice to the human rights worker, the disenfranchised, the oppressed, the builder, and the whistleblower. They also give voice to the troll, the racist, the enemy state, the bot, and the bot army. Unlike traditional media that hired journalists and broadcasters, modern platforms do not create the content they distribute. Both by law and by design, they absolve themselves of responsibility for propagating fake news, even as they expand the population capable of spreading it. A third reason is false reality born of "deep fake" technology. Victims of political slander could once credibly deny as hearsay false and defamatory claims by third parties. Deep fake technology, however, can create first person fictions that make a victim appear to have committed acts he or she never did or spoken words he or she never would. In appearance and behavior, forged details have become indistinguishable from originals. And, still a fourth reason is the potential for systemic failure. A group of ecologists, biologists, computer scientists, and political scientists warns that social systems, like biological systems, can exhibit cascade failures when pushed too far.[8] Like coral ecosystems that have withstood millennia yet passed in an ecological instant, social systems can break down when pressed beyond their institutional constraints and citizens cannot agree on basic facts.

---

[4] "Inside America's Shocking WWII Propaganda Machine". December 19, 2016. Retrieved August 19, 2018.
[5] [Wikipedia 55, 57, 58]xxx.
[6] Trump campaign introduced disappearing ads on Facebook in Wisconsin, North Carolina and Georgia to deter black votes. Channel 4 News https://www.youtube.com/watch?v=KIf5ELaOjOk
[7] The counter argument that Clinton could have bought similar ads is too simplistic. Without knowing what has been said to whom, the cost of covering all possible arguments among all possible listeners is prohibitive. The chief beneficiary, in this case, would be the platform selling blanket advertising.
[8] Bak-Coleman, J. B., Alfano, M., Barfuss, W., Bergstrom, C. T., Centeno, M. A., Couzin, I. D., ... & Weber, E. U. (2021). Stewardship of global collective behavior. *Proceedings of the National Academy of Sciences*, 118(27).

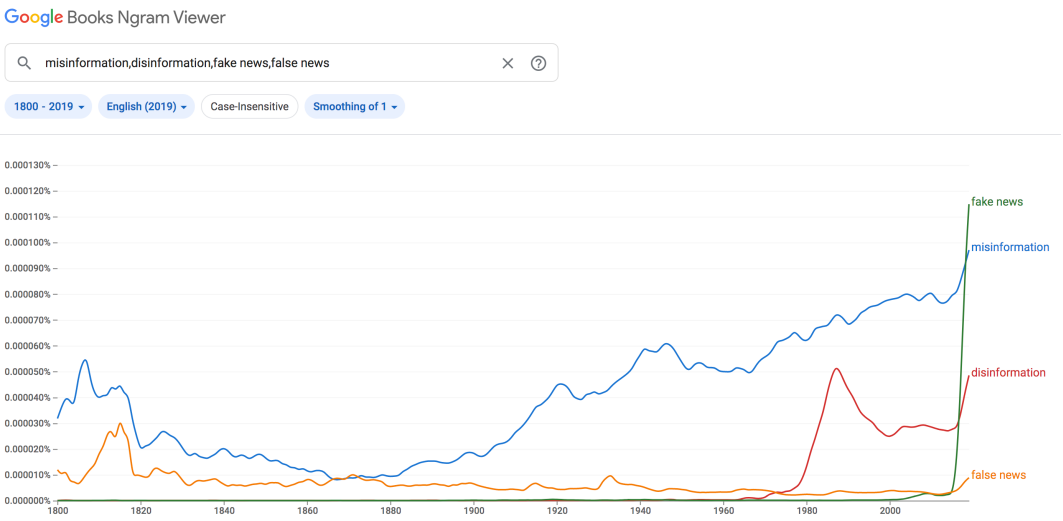

Figure 1 – Occurrence of fake news terminology in books and literature 1800-2019. Other terms dominated the term "fake news" until 2016. A rise in misinformation occurred around World Wars I and II, a rise in disinformation occurred during the Cold War, and in fake news in 2016.[9]

Fake news is a hard problem. Open societies tend toward higher growth rates,[10] greater freedom of expression, and greater justice[11] than closed societies. Censorship is a preferred tool of despots and dictators. Freedom of speech is a fundamental right precisely because it exposes those with much to hide. At the same time, a right to free speech must balance a right to privacy and a right to self-rule. Not every fact of one's private life deserves public scrutiny. No lie justifies overturning a fair election. Yet, it can also happen that ugly truths can be more divisive than harmless lies. Design of a fair and balanced news distribution mechanism must weigh rights and properties that conflict with one another yet promote the public good. Presumably, the balance of these rights should be robust to circumstance and not the fragile reactions to politics of the times.

The problem of rights is a problem of balance. Freedom of expression can be used to harass others for exercising their freedom of expression. One right obliterates another. Freedom of expression as a virtue has obvious bounds as a vice.

This article seeks to do three things. The first is to articulate why fake news is a problem and how harm occurs. From this understanding, the second is to lay a foundation for a modern framework of addressing it. Social media platforms of the 21st century differ in material ways from the print and broadcast media of the 20th. The third is to articulate tests by which one such mechanism might be compared with another. There is no promise that such a framework is either correct or complete, only that it is as balanced as this author can make it.

---

[9] Source: http://books.google.com/ngrams

[10] Przeworski, A., Limongi, F., & Giner, S. (1995). Political regimes and economic growth. In *Democracy and Development* (pp. 3-27). Palgrave Macmillan, London.

[11] Acemoglu, D., Johnson, S., & Robinson, J. A. (2005). Institutions as a fundamental cause of long-run growth. *Handbook of economic growth*, *1*, 385-472.

Definitions: What is Fake News

A number of scholars have offered a variety of fake news definitions. From their descriptions, a brief summary includes:

- Fraudulent high velocity content[12]
- News articles that are intentionally and verifiably false, and could mislead readers. … It also includes many articles that originate on satirical websites but could be misunderstood … when viewed in isolation[13]
- The online publication of intentionally or knowingly false statements of fact … is a complex and fact-specific endeavor better addressed through case-by-case analysis.[14]
- … is determined by fraudulent content in news format and its velocity.[15]
- "stories that are provably false, have enormous traction [popular appeal] in the culture, and are consumed by millions of people"[16]
- Fake news is written and published with the intent to mislead in order to damage an agency, entity, or person, and/or gain financially or politically,[17] often using sensationalist, dishonest, or outright fabricated headlines to increase readership, online sharing, and Internet click revenue.[18]

Some scholars in this list assert that author intent matters. Satire, parody, and entertainment, without deceitful intent, do not constitute fake news according to this view. Interestingly, the 1938 broadcast by H.G. Wells of a fake Martian invasion would then not qualify as fake news, despite panic due to belief in its authenticity, as Wells' purpose was entertainment not deceit. Clarifying nomenclature distinguishes "misinformation," which can include "inadvertent sharing of false information" versus "disinformation," which refers to "deliberate creation and sharing of information known to be false".[19]

---

[12] World Trends in Freedom of Expression and Media Development Global Report 2017/2018. http://www.unesco.org/ulis/cgi-bin/ulis.pl?catno=261065&set=005B2B7D1D_3_314&gp=1&lin=1&ll=1: UNESCO. 2018. p. 202.

[13] Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. Journal of economic perspectives, 31(2), p. 213.

[14] Klein, D. & Wueller, J. "Fake News: A Legal Perspective" Journal of Internet Law 20(10) April 2017. P. 6.

[15] World Trends in Freedom of Expression and Media Development Global Report 2017/2018 UNESCO p 202

[16] Michael Radutzky, producer of CBS *60 Minutes*.

[17] Hunt, Elle. "What is fake news? How to spot it and what you can do to stop it". The Guardian; Dec. 17, 2016.

[18] "The Real Story of Fake News" Merriam Webster https://www.merriam-webster.com/words-at-play/the-real-story-of-fake-news.

[19] Wardle, C. (2017). Fake news. It's complicated. *First Draft News*, *16*. https://medium.com/1st-draft/fake-news-its-complicated-d0f773766c79

Tandoc, Zeng & Ling[20] add the role of the audience, further conditioning whether news is fake on the listener's belief as distinct from the author's intent. They argue that without deception, fake news is but a work of fiction. Interrogating subjective and unobservable factors such as author intent and listener belief are challenging but suggest the use of mechanism design to handle information asymmetry. "News you don't believe" is a murky colloquial[21] shortcut. If it is fake news that is disbelieved, then it does no damage and requires no intervention. If it is true news that is disbelieved, then it is not fake but could do damage, an irony that presents design challenges.

To put it succinctly, the challenge of designing a fake news intervention based on the forgoing attributes appears to hinge on addressing veracity (falsifiability of news), velocity (speed and reach of news dissemination), and volume (amount and frequency of news production). It may also need to address author intent and listener belief.

Spreading False Information Need Not Cause Harm

Textbooks prior to 2006 published the existence of nine planets. In that year, the International Astronomical Union reclassified Pluto as a dwarf planet,[22] following the 2005 discovery of Eris[23] a celestial body 27% more massive than Pluto. If the larger body was not a planet, then how could the smaller body be one? Texts prior to 1781 suggested there were only six planets.[24] Neither the six nor nine planet claim was true[25] yet the lives of few individuals changed in any meaningful way as new facts corrected old falsehoods.

One Asian fusion restaurant in Cambridge, Massachusetts used the slogan "Eat at Jae's and live forever!"[26] The National Enquirer, a publication with a circulation that reached one

---

[20] Tandoc Jr, E. C., Lim, Z. W., & Ling, R. (2018). Defining "fake news" A typology of scholarly definitions. *Digital journalism*, *6*(2), 137-153.

[21] Nielsen, R. K., & Graves, L. (2017). News you don't believe": Audience perspectives on fake news. *Reuters Institute for the Study of Journalism. Retrieved from https://reutersinstitute. politics. ox. ac. uk/ourresearch/news-you-dont-believe-audience-perspectives-fake-news*.

[22] Resolution B5: Definition of a Planet in Our Solar System. https://www.iau.org/static/resolutions/Resolution_GA26-5-6.pdf Accessed August 19, 2018.

[23] "Eris is the Greek god of discord and strife, a name which the discoverer Mike Brown found fitting in the light of the academic commotion that followed its discovery." Pluto and the Developing Landscape of Our Solar System. International Astronomical Union https://www.iau.org/public/themes/pluto/. Accessed August 19, 2018.

[24] Uranus was discovered March 13, 1781 by William Herschel. Neptune was discovered Sept 23, 1846 by Urbain Le Verrier, Johann Gottfried Galle, and John Couch Adams. Pluto was discovered Feb 18, 1930 by Clyde Tombaugh

[25] Logically, as long as the definition is consistent, the number cannot be six or nine. If the definition includes dwarves, then the set of planets includes previously undiscovered bodies Eris and Ceres for a set of at least eleven. If one excludes dwarf planets, then the set includes only eight.

[26] See Boston Globe review posted http://archive.boston.com/dining/globe_review/1087/. Jae's discontinued using the slogan after several locations closed.

million, published a cover story that actress Rita Hayworth had returned from the dead.[27] Each one of these claims is provably false. Few of these claims have altered decisions resulting in consequential damages. None of them, even in total aggregate, have produced social costs rising to a level that would require regulatory oversight.

Even the belief in false news need not cause harm.[28] One might believe that vaccines do not work yet still take them in compliance with the law or one might believe the world is flat yet still take cruises that sail the globe. Those beliefs had no ill effect. People have built cathedrals, written songs, sculpted art in the belief that a martyr had risen from the dead.[29] Only when one acts on beliefs that cause harm does fake news lead to a social interest in curbing the harm. Then also, restricting actions can occur without revising beliefs; they represent separate points of intervention with different ethical and efficiency considerations.[30]

In fact, lies can even create value. Members of the underground railroad, who lied to bounty hunters after passage of the Fugitive Slave Acts of 1793 and 1850, faced fines and imprisonment but helped thousands of refugee slaves escape to freedom.[31] During the holocaust, villagers of the town Vivarais-Lignon lied to Gestapo officers about the presence of Jewish refugees, risking their own execution, yet they helped people in need evade Nazi concentration camps.[32] Collectively, such lies saved hundreds if not thousands.

Falsity alone is not a metric that can determine the need for intervention.

Spreading True Information *Can* Cause Harm

During World War I, the English language North China Daily News printed allegations that a German factory was rendering human corpses into fats to produce nitroglycerine and lubricants.[33] This story was false and intended to gain allies in the war. During World War II, Nazi Propagandist Joseph Goebbels used the truth about these stories to discredit other

---

[27] "I'm back from the dead – For two years I was a zombie" *National Enquirer*; Vol 35, No. 14. Dec. 15, 1963

[28] Harari, Y. N. (2018). *21 Lessons for the 21st Century*. Random House, ch 17.

[29] Ibid.

[30] What are the intervention considerations for using which? Likely scale of light touch efficiency of limiting action, next level is limiting spread of false beliefs (and limiting action), top level revising acts and false beliefs (and limiting action). False beliefs that cause neither decision error nor externalities do not require intervention. Harm (via either decision error or negative externalities) requires intervention regardless of truth or falsity of news. When does a person need to be protected from self-deception? We don't allow lies about medical products leading to belief in efficacy.

[31] https://www.history.com/topics/black-history/fugitive-slave-acts

[32] https://time.com/5680342/french-village-rescued-jews/

[33] British Broadcasting Company, "The Corpse Factory and the Birth of Fake News." 17 February 2017. https://www.bbc.co.uk/news/entertainment-arts-38995205

true stories of German war crimes against the Jews. In effect, he used truthful news to discredit truthful news.[34]

> *"To tell a truth with ill intent beats all the lies you can invent."*
>
> *William Blake 1919*

In 2009, the largest data breach in history took place when a lone hacker broke into Heartland Payment systems and exposed the data of 130 million credit card holders. The practice, common among thieves, is to monetize either directly by billing these accounts or indirectly by selling them to other criminals. The more accurate and complete the data, the more value it has on the black market and, correspondingly, the greater is the damage to card holders.

In 2018, a US federal court banned the free dissemination of blueprints for 3D printed guns. If the blueprints had been inaccurate or fake, they could not have been used to print working guns. The judge banned them on the basis that harm to the private defendant's First Amendment rights "are dwarfed" by the harm States might incur if anyone could print a gun.[35] Criminals could defeat security scanners. Printed plastics would facilitate terrorist hijackings. Persons legally barred from gun ownership, due to prior conviction, restraining order, or mental condition could summarily obtain them.

In 2016, Russian troll accounts Blacktivist and DrConservaMom used social media to broadcast true stories of white officers shooting black men and of school shootings. Their messages used true information, tinged with political spin, to suppress black votes in neighborhoods favoring democrats and to animate gun rights voters in neighborhoods favoring republicans. Twitter identified more than 50,000 accounts linked to Russia,[36] and suffered a 21% stock loss after purging more than 1 million fake accounts.[37] Facebook claims have deactivated numerous fake accounts prior to the US 2020 election.[38] It is easy to cull such messages on the premise of foreign interference in sovereign elections. It is not

[34] Marlin, Randal (2013) **Propaganda and the Ethics of Persuasion**. Broadview Press.
[35] Judge Robert. S. Lasnik: Washington v. US State Department NO. C18-1115RSL July 31, 2018.
[36] Koh, Y. "Twitter Reveals 1,000 More Accounts Tied to Russian Propaganda Agency," *Wall Street Journal*. Jan 22, 2018.
[37] Market Watch, "Twitter stock plunges 21% after earnings show effects of fake-account purge"July 27, 2018. https://www.marketwatch.com/story/twitter-shares-slide-16-after-fake-account-purge-new-rules-in-europe-2018-07-27
[38] https://www.cnbc.com/2019/11/13/facebook-removed-3point2-billion-fake-accounts-between-apr-and-sept.html

so easy to cull such messages propagated by domestic citizens themselves. Indeed, multiple such accounts were used in the 2016 and 2020 US elections.[39]

Among ideologues and propagandists, it is common practice to take two truths and falsely link them. A famous celebrity who did die and the fact that he did receive a vaccine a few weeks earlier are used to discredit vaccinations that had nothing to do with his cause of death.[40] Antivaxxers routinely cite a small piece of evidence from legitimate research, remove context, and grossly exaggerate it.[41]

Veracity alone is not a metric that can determine the need for intervention.

Of these truthful news examples that do cause harm, each represents a negative *externality*. In the case, of disclosed credit cards, the harm accrues not to the repository or to the thief. Harm accrues to a third party – the card holders whose private data was misappropriated. In the case of Goebbel's misuse of true information, the harm was not to Goebbels or readers of his fake news but to the Jews who suffered historic atrocities. In the case of working instructions for printed guns, the harm is not to the author of blueprints or the consumer who prints them but to innocent victims who are shot using them. In the case of voter suppression, it is not just the citizens who don't vote but all other citizens who become governed by a different choice of candidate. In the case of antivaccination exaggerations, it is not just the unvaccinated themselves but also the failure of herd immunity and economic costs to the community.

Beyond the issue of veracity, fake news constitutes a form of information pollution from which there extend potentially large externalities. The need to address externalities as well as veracity leads to the following definition of the problem.

The phenomenon that this article will address is to clear a communications channel of dysfunctional information i.e. that which causes net social harm when propagated at scale. The chief information characteristic causing harm is causing either decision error or negative externalities among a population and not merely individuals.

This problem statement differs from prior interpretations in two important ways. First, decision error rather than truth per se renders the task easier and separates the problem from intent, which may or may not be discernable. It also identifies half-truths that deceive listeners into choosing differently than how they would have chosen under full information.[42] Intent might matter at a penalty phase but is rarely available at the decision

---

[39] Aral, S. (2020). *The Hype Machine: How Social Media Disrupts Our Elections, Our Economy, and Our Health--and How We Must Adapt*. Currency. https://www.bloomberg.com/news/articles/2020-09-01/social-media-impersonators-seen-as-threat-in-upcoming-elections

[40] https://www.npr.org/2021/05/13/996570855/disinformation-dozen-test-facebooks-twitters-ability-to-curb-vaccine-hoaxes (ibid)

[41] Ibid

[43] Testimony before the Joint Commerce and Judiciary Committees, April 10, 1918.

to disseminate phase. Second, it adds a missing component, harm to third parties. A solution that only envisions harm to the listener fundamentally misses the negative externalities that harm others. In his testimony before Congress describing Facebook's interaction with Cambridge Analytica, Zuckerberg stated "We did not take a broad enough view of our responsibility."[43] Cambridge Analytica promised games and surveys to Facebook users but in exchange harvested not only all their individual data but also their friends' data, a textbook example of a third-party externality. Friends did not grant permission for data harvesting and were unaware of the breach.[44] The spillover consequence for others is precisely the issue of third-party harm.

Solutions

I – The presence of an externality means that knowledge of the transaction is divorced from knowledge of the harm. The information sets do not overlap. A framework for addressing the problem then admits one of two solutions. The first option is a governance mechanism that moves information about the harm to the party with knowledge of the transaction. Facebook, for example, could learn of damage that its ads cause. The second is a governance mechanism that moves information about the transaction to the party with knowledge of the harm. People affected by ads on Facebook, for example, could learn details of the ads.

The first option is inferior for at least three reasons. (i) Concentrating all information at the center creates a platform of large power and little oversight. At the same time, concentrating off-platform information on-platform has the potential to increase information asymmetry across society creating the potential for widespread exploitation. (ii) Moving information from off-platform to on-platform does not align incentives. Since the party with knowledge of the transaction is *not* the party suffering harm, the central platform need not be motivated to change behaviors so as to improve social welfare. (iii) The near infinite variety of potential externalities ensures that certain forms of damage are likely to be missed. Pulling all possible information with all possible externalities onto the platform is a technical task so daunting that, given any private information among individuals, obtaining complete knowledge is nearly impossible. In effect, the reasons that Facebook "did not take a broad enough view of [its] responsibility" include both that it is technically infeasible and incentive incompatible. Moreover, were it possible to succeed, the outcome would not be desirable.

The second option is superior for reasons that invert the logic above. (i) Moving information that is on-platform to third parties off-platform decentralizes power and reduces information asymmetry. The prospect of exploitation falls. (ii) Parties that suffer harm obtain information

[43] Testimony before the Joint Commerce and Judiciary Committees, April 10, 1918.
[44] Cadwalladr, C. (2018). I created Steve Bannon's psychological warfare tool': Meet the data war whistleblower. *The Guardian*, *17*. https://www.theguardian.com/news/2018/mar/17/data-war-whistleblower-christopher-wylie-faceook-nix-bannon-trump

on the causes of harm. Incentives align, giving those with the desire to act the information needed to act. Welfare naturally improves. (iii) Public exposure, moving on-platform information off-platform, facilitates parallel search by diverse members of a society. This is decentralized rather than centralized governance. The chances for uncovering the nature of harm improve. At the same time, decentralization fosters a marketplace of information where different ideas compete. Truthful interpretation is easier on the basis of shared facts.

1) Access Solution: Pair knowledge of harmful externality with access to means of resolution – grant "equivalence of access"

This is not merely a matter of transparency or maintaining records for a period of years and making the nature of an ad, advertiser, and contact information available. It must also provide access to those who saw the ad and to those where it spread. This was infeasible in print and broadcast media. This provides means to *undo* the damage, not merely allow for its discovery. Information must be actionable not merely knowable. In effect, this rule resembles the "Equal Time Rule" that requires broadcast stations to provide equivalent access to present their case if requested. On balance, this solution should largely be business model compatible as it simply allows one partisan to use the same advertising tools as another partisan.

Resolving information asymmetry between those doing the harm and the locus of harm suggests a first intervention. Media platforms should provide access to those who have received dysfunctional information. That is, media platform should grant an injured party access to media targets equivalent to that obtained by the parties causing injury. Access should include, for example, not only people reached by an ad but also those contacts to whom such ads were shared. Access goes beyond current transparency requirements that record who purchased an ad, its content, and release dates. Transparency only lets affected parties learn that damage has occurred. It does not provide means of undoing the damage. Limited disclosure only provides information sufficient to reach the perpetrator, not the means to undo harm by enabling counter messaging. By contrast equivalence of access allows an injured party to seek redress not only by holding the perpetrator accountable, which transparency provides, but also by updating facts and narratives among recipients, which access provides.

Transparency laws that simply record the nature of a communication – its sponsor and its content – do not per se address decision errors or externalities. If a crime were committed, such laws provide only for recording the event or compensating it without undoing the damage. Transparency allows people to discover how they were injured and to hold the speaker accountable. They do not, however, enable the injured party to reach those members of the market for ideas where the false idea has taken root. Transparency might allow one to sue a candidate for a lie placed prior to an election but access can allow one to undo the lie that changed the outcome of that election. If awareness of an externality is insufficient to correct it, the access to means must accompany knowledge of ends. Media

platforms must therefore record recipients of ads received directly as well as those who received them indirectly via sharing and promotion.

Market access need not affect privacy of individual recipients. Neither their identities nor their contact information need be disclosed. Instead, the platform simply mediates access via ads.

Importantly, media platforms should find provision of equivalent market access to be business model compatible. They need not arbitrate truth. They simply sell ads with access to all parties equally. In effect, "Equivalence of Access" on social media resembles the existing FCC "Equal Time Rule" that requires broadcast stations to provide equivalent market access to present their case if requested. "Equivalence of Access" is both distinct from and weaker than the "right of reply" that required free placement of rebuttals on behalf of citizens disparaged in broadcast editorials. The issue in Red Lion, a right of reply attached to broadcaster editorializing, hinged on the limited bandwidth. A person who felt attacked by a radio station had few ways to reply and launching a competing radio station was impractical. A legally binding free response might also cause courts to intervene if the broadcaster felt an editorial did not merit a response. Government then needed to adjudicate content. By contrast, this attaches to third parties off-platform. In no case does equivalence of access require the media platform to take a position or speak in a different voice. It only provides a channel for access at published and prevailing rates. As with access to patent pools, terms should be Fair Reasonable And Non-Discriminatory (FRAND). It balances the media platform's interest in protecting its assets and the social concern with fair markets for ideas. Government plays no role in adjudicating content and, if involved, only decides whether market access is fair.

A right to reply was weakened by the Supreme Court in Miami Herald Publishing v Tornillo 418 U.S. 241 (1974) in the case of newspapers. While a law requiring a right of reply did not prevent editors from saying what they wished, "it exacts a penalty on the basis of the content" and because newspapers are financially limited, "editors may conclude the safe course is to avoid controversy." By contrast, equivalence of access boosts revenues and, if anything, invites crosstalk among political opponents. Ironically, equivalent access, paid at FRAND rates, has the opposite effect of the court's concern for an economic burden on the press. Equal access at FRAND rates introduces the moral hazard that a media platform could invite or offer critique in an effort to prompt those affected to purchase a response. It is certainly business model compatible. The truth of each perspective could then surely enter the market.

A partisan might object that a media platform could fill all ad space with lopsided content and thereby protect association with a particular point of view. This might be acceptable for a citizen group that is self-funded for advocacy. It is not acceptable for a group whose function is, in large part, to reach an audience. Such a partisan objection does not apply, as in this case, to platforms funded by advertising. While laudable, the goal of protecting a uniform identity must be deeply subordinated to the goal of providing access to the market

of ideas. No society should tolerate discrimination against citizens' voices on the basis of race or gender. The same is also true for points of view. Transparency resolves information asymmetry regarding knowledge of harm. Equality of access permits the undoing of harms.

One shortcoming of this solution is granularity. It applies at the ad or institutional level rather than the individual level. Transparency of messaging is less feasible and less desirable at the individual connection level. An intervention motivating individual care in sharing higher quality news appears next.

II – One clear cause of the information pollution problem is that lies spread "farther, faster, deeper and more broadly than truth in all categories of information."[45] Platforms spread a wildfire of lies in order to build businesses based on engagement. If their goal is to attract eyeballs, then flames will do the trick.

Platforms promote blowhards in the name of newsworthiness. A Cornell study of 38 million articles found that the single greatest source of coronavirus misinformation – including that disinfectant is a cure, that an anti-malaria drug is a cure,[46] that the pandemic was a democratic party hoax, and that masks are not effective at reducing viral spread – was the US president.[47] False claims undermining integrity of the U.S. 2020 election reached the point that they led to insurrection.[48] Reversing this amplification suggests a second solution.

2) Friction Solution: Add social friction to liars and not just their lies

A straightforward way to implement a "social friction" policy is to selectively reverse platform amplification. Platforms could adopt a policy that parties convicted of lying will have their social networks trimmed and their messages delayed. Does a person have 100,000 followers? Following a lie, it will be 50,000 and messages will go out a week later. A badge of dishonor, applied to the liar, can inform followers why the platform no longer pushes that liar's messages into followers' news feeds. Penalties can apply temporarily for good behavior but increase for bad behavior. Repeated lying could means having 25,000 followers and messages go out every two weeks. Liars can still say what they wish, even to the point of lying, but then followers would need to go looking for their misinformation in contrast to having the platform promote it.

---

[45] https://science.sciencemag.org/content/359/6380/1146.full

[46] This was never seriously in doubt. Malaria is caused by a unicellular parasite whereas covid-19 influenza is caused by a virus. Cellular organisms and viruses operate in different ways. Unsurprisingly, antiviral drugs are more effective than antiparasitic drugs in treating covid-19 https://www.cidrap.umn.edu/news-perspective/2020/10/new-covid-studies-remdesivir-yes-hydroxychloroquine-no

[47] https://allianceforscience.cornell.edu/wp-content/uploads/2020/10/Evanega-et-al-Coronavirus-misinformation-submitted_07_23_20.pdf

[48] https://www.usatoday.com/story/opinion/2021/01/13/trump-disinformation-campaign-led-to-capitol-coup-attempt-column/6639309002/

Social friction has three benefits. First, it directly addresses the problem that lies spread faster, farther and more broadly than truth. It filters pollution at the source. It reduces and delays the channels through which lies spread. Second, social friction motivates liars to change their behavior. If the goal of a blowhard or ideologue is to attract attention or to move an audience, what better motivator is there than limiting audience access? Labeling and deleting do not work. Unmotivated by truth or integrity, they rephrase and repost. Undeterred and without penalties, ideologues with large networks volley and amplify each other's false claims.[49] When social friction applies to liars, ideologues choose to shrink their own audiences by telling lies. Networks of echo chambers that willfully propagate lies then self-destruct as they willfully take themselves down. What we have needed is a mechanism that disproportionately weeds out untruths as compared to truths when, up to now, we have had the opposite. Going forward liars render themselves less potent by limiting their own reach.

The incentive structure highlights a third benefit: This policy places the burden on the proper source, the liar rather than the platform or the reader. Too many attempts at solutions insist that platforms mediate 500 million daily messages[50] that they do not author – do we want them judging every message? Can they? Or, proposals ask readers to sift through mountains of manure to find the truth – will they? Who knows better that a claim is false than the author of that claim? Putting social friction on liars causes authors to think and deliberate before pushing what they know to be false.

Friction does not eliminate false information or ability to speak. Rather, it shifts from all-or-nothing censorship, where information is lost, to a graduated increase in difficulty of dissemination, where information is retained. This is especially useful to society if, in some future condition, the purported falsehood turns out to have been true. The datum is discoverable and the processes used to vet it can be improved.

The economics do not favor platforms voluntarily adding friction. This solution is not intrinsically business model compatible. It is cheaper to produce fake news than true news. By enfranchising everyone, social media platforms shift the balance of supply toward cheaper and therefore more abundant sources of supply. The volume of fake news increases. At the same time, fake news is more engaging. It spreads faster, farther, and more broadly than truth.[51] As demand is higher for novelty, machine learning algorithms push that which generates engagement. Social media embrace this demand. Across production and consumption, a population's news diet shifts. The business model optimizes

[49] Zakrzewski, C. "Trump's Twitter feed is covered in warning labels," Washington Post. Nov. 5, 2020. https://www.washingtonpost.com/politics/2020/11/05/technology-202-trump-twitter-feed-is-covered-warning-labels/

[50] https://www.internetlivestats.com/twitter-statistics/

[51] Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146-1151. (Ibid.) https://science.sciencemag.org/content/359/6380/1146.full

for profit orthogonal to truth, human health, and institutional health.[52] Senators have chastised such platforms for taking insufficient action against only a dozen individuals that spread up to 65% of vaccine misinformation.[53] Platforms have eschewed reducing politicians' fake news on the basis it is newsworthy.[54] Adding friction reduces engagement which reduces profits.

III – Addressing the fact that platforms are not well-motivated to self-correct their problems prompts discussion of regulatory solutions for decision errors and externalities. At present, we know of only two solutions for solving externality problems. The first, proposed by Arthur Pigou, levies a tax proportional to the damage in order that the marginal private cost rises to the marginal social cost. The producer then internalizes the harms rather than shifting them to the citizenry. The second, proposed by Ronald Coase, creates property rights in the externality and uses markets to trade and price it.[55] This raises the cost of otherwise free disposal, while also shifting the burden of harm to whomever can bear it most cheaply. We develop each option in turn.

3) Pigouvian Solution: Tax the platform in proportion to the harm

At least one Nobel Economist has endorsed a variant of the Pigouvian tax intended to focus on the business model rather than the externality per se.[56] Romer suggests applying a Pigouvian tax to digital ad sales. This offers at least three advantages over alternative interventions. First, an ad tax directly alters the business model by favoring ad-free subscription revenue over ad-based third-party revenue. Subscription revenue need not require user tracking. Also, above a minimum participation threshold, subscription revenue does not create an incentive to artificially boost engagement. User privacy could improve even as fake news driven demand declined. Second, a progressive tax, with higher costs for larger firms, could favor entrepreneurial startup. Larger firms created by smaller firm mergers would face larger ad tax bills. Relative to breakup, a progressive ad tax simultaneously solves a market concentration problem normally solved through competition law but does so more effectively. Two half size firms could produce more damage than one full size firm if the principal effect of breakup were to cause each smaller firm to compete more fiercely for user engagement. Not so for the ad tax. Third, an ad tax is not content based, which avoids all free speech concerns. It removes government from adjudicating speech. And, because it does not involve operational oversight, it reduces risk of regulatory capture that can arise from oversight. Both the Federal Aviation Administration

---

[52] Bak-Coleman, J. B., Alfano, M., Barfuss, W., Bergstrom, C. T., Centeno, M. A., Couzin, I. D., ... & Weber, E. U. (2021). Stewardship of global collective behavior. *Proceedings of the National Academy of Sciences*, 118(27).

[53] https://www.npr.org/2021/05/13/996570855/disinformation-dozen-test-facebooks-twitters-ability-to-curb-vaccine-hoaxes

[54] https://www.nytimes.com/2019/10/17/business/zuckerberg-facebook-free-speech.html (ibid)

[55] Hovenkamp, H. (2009). The Coase Theorem and Arthur Cecil Pigou. Ariz. L. Rev., 51, 633.

[56] Romer, P. "A Tax That Could Fix Big Tech," New York Times; May 7, 2019; Section A, p. 23. https://www.nytimes.com/2019/05/06/opinion/tax-facebook-google.html

and Interstate Commerce Commission have faced charges of inappropriate ties to the aviation and trucking industries respectively.

Romer's version of a Pigouvian tax has two shortcomings. First, the primary means by which all platforms create value is by consummating matches.[57] They pair people with friends, news, apps, search results, rides, movies, products, and destinations. Effective matching requires tracking. Purported privacy benefits will not fully materialize although subscription revenues do align the interests of user and payer in a way that ad revenues do not. Second, one of its greatest strengths is also its greatest weakness. The damage targeted by an ad tax is unhealthy levels of engagement not fake news per se. By avoiding content issues, the ad tax divorces the levy from the externality. A private subscription service could host antivaxx disinformation, conspiracy theories, and false election narratives but pay no tax, whereas a clean ad-driven service, free of fake news, could pay a heavy tax. For Pigou's solution to work, the penalty must scale with the externality. In the context of factory pollution, a tax on the percentage of harmful effluent encourages the factory to shift technology but only if it corresponds to the harm. By contrast, if the tax applies to all output regardless of harm, the factory keeps using the lower cost more polluting technology.

This insight offers a means of reforming a progressive ad tax to improve efficacy: tie the levy to the concentration of harms produced in news effluent. This functions exactly in the manner of taxing the concentration of harmful effluent in factory production. Yet, a tax on falsehood spillovers acts differently than a tax on specific speech. Testing a statistically valid sample solves three fundamental problems, one of scale, one of accuracy, and one of law. First, one need not certify every message; rather a certification authority need only validate a random sample in order to achieve any confidence level desired. Sampling could even apply to closed chat rooms without violating the privacy of the individuals involved. Second, in a rigorous mathematical sense, a flow rate or aggregation of signals provides a constantly updating Bayesian credibility score. Based on the central limit theorem, larger samples cause estimates of any parameter to converge closer to truth as samples accumulate. This advantage is enormous as it deals even with mixed stories that blend truth with lies. An overarching news credibility score characterizes fitness with respect to the whole environment and not simply a single event. Parties on the left and right might disagree on which messages are true yet agree more readily on the flow rate of truth. Given sufficient statistical samples, consistent deviation from an average score can indicate bias in a specific critic as easily as bias in a specific critique. Lastly, a third benefit is that a tax on concentration of harm offers a practical means to weaken the liability protections of Section 230 while retaining its broader benefits. The binary choice between either total liability immunity or accountability for all individual messages is too coarse. Platforms can reasonably object that policing content of 500M individual messages daily is not practical. Societies can reasonably object that policing disinformation that causes political insurrections, unnecessary deaths from infections, and genocidal riots had better be

[57] Parker, G. G., Van Alstyne, M. W., & Choudary, S. P. (2016). *Platform revolution: How networked markets are transforming the economy and how to make them work for you*. WW Norton & Company.

practical or the platform should not operate. A reasonable balance – one that can adapt to different societies – is to hold platforms accountable for a specific fraction of effluent. Tax the preponderance of dysfunctional information as distinct from specific instances of disinformation. Measuring a fraction of effluent can exhibit practical scale, adapt to law, and converge to truth.

IV – Truth in markets for political ads is particularly troubling. At one extreme, Facebook defended its decision to take all ads, including outright lies, on the basis that citizens, not Facebook, should decide.[58] At the other extreme, Twitter chose to ban all political ads to avoid both lies and bias.[59] The former pollutes our discourse with fake news unfiltered. The latter entrenches incumbents and impoverishes discourse. There is a better solution. By combining Coase's ideas on externality markets and information economics signaling, we can create a "market for truth." It requires neither machine algorithms to discern truth nor judgments by a potentially self-interested company. Instead, it discourages liars from lying. It would work as follows:

4) Extended Coasian Solution: Offer ad guarantees in a market for truth

Give people making strong claims the option (not a requirement) to warrant that their claims are true. Examples include a politician making a claim about an opponent, a policy officer or antivaxxer making a claim about vaccine efficacy, or a consumer product company making a claim about its product efficacy or where it is made (e.g. "made in the USA"). The warrant, posted in advance, serves as a time-limited reward to anyone who can prove the claim is false. To dispute a claim, a challenger pays a modest fee to cover the cost of adjudicating fact-checking. Adjudication is handled by a random sample of peers. Winning challengers claim the warranty to spend as they wish, allowing them to undo the harmful externality. Unchallenged claims, or those judged true, have the warrant returned to the author.

In all cases, the cost of guaranteeing the truth of an honest ad is zero. The false claimant, however, pays for the ad, pays the pledge penalty, and pays in reputation. Simply put, the forfeited pledge is the price of a lie. It is paid only by liars. A politician who wishes to lie may still do so. But lying becomes expensive.

What about the slippery middle ground between truth and falsehood— the innuendo and half-truths that infect so much political advertising? Imagine a photo of Joe Biden and his son looking shifty, accompanied by the tagline: "Hunter Biden served on the board of Ukraine's most corrupt company while his father, as Vice President, did all he could to fire a powerful Ukrainian prosecutor." None of that is exactly false. But it implies the senior

[58] https://www.washingtonpost.com/technology/2019/10/17/zuckerberg-standing-voice-free-expression/
[59] Twitter banned "content that references a candidate, political party, elected or appointed government official, election, referendum, ballot measure, legislation, regulation, directive, or judicial outcome." https://www.vox.com/recode/2019/11/15/20966908/twitter-political-ad-ban-policies-issue-ads-jack-dorsey

Biden tried to prevent the prosecutor from going after the company, when in fact he sought the opposite: he wanted the prosecutor fired for failing to pursue corruption.

How should an honest ads market handle an ad like this? It refunds half the pledge for an ad that's half a truth. Based on the egregiousness of the lie, the amount of a refund can correspond to one of the sliding scales fact checkers already use. Indeed, Politifact did rate an ad like the one here as half-true on a scale that ranges from: true, mostly-true, half-true, mostly- false, false, and pants-on-fire. Other fact checkers use similar scales. A market for truth need not be perfect. It just needs to be credible and unbiased. By allowing PACs and politicians to warrant their claims, it changes the balance of power, favoring truth over lies in our political discourse.

In what sense is this a Coasian solution? Why would this work? The extended solution combines externality economics with information economics. A truth market for trading honest ads works for the same reason as a carbon market based on cap and trade. It solves the problem of pricing externalities and markets for trade in externalities already exist. Carbon dioxide is pollution. It is a negative externality that harms others. An entity that is causing damage must pay for that damage by buying pollution credits that put a price on the harm done. Fake news is pollution. It is a negative externality that harms others. The size of the honest ads pledge, that is, the lie price, could be any escrow amount set by the social media platform but should be the expected size of harm done. This negative externality is the "social cost" of the damage done by lying. The crowdsource identification of harm is the market that "trades" the externality. The harmed parties claim the lie price and get paid for the damage they experience. Carbon trading markets work so we can expect markets for truth will also work.

Importantly, a market for truth works even when the amount of damage, the lie price, is not known in advance. Imagine Exxon Mobile today taking out an ad that human activity does not cause global warming. The lie price for political ads in the U.S. alone is too small for the lie price of global warming policy internationally. One can quickly see that, if a firm repeatedly pays the lie price, then their willingness to keep lying is too small relative to the true social cost. Then the lie price should rise until they stop the lies that cause harm. In other words, we have an "efficient search" process that can force firms and super PACs to internalize the true social cost of their negative externalities even when that cost is initially unknown.

And what about free speech? In the U.S., skeptics might object that an honest ads pledge would not withstand First Amendment scrutiny if the pledge were mandatory. U.S. courts view impediments to speech as violations of free speech. Although this is a uniquely U.S. problem, the system still works even when a pledge is voluntary. If the market for truth is fully functioning, then unwillingness to pledge an honest ad is itself a signal that the author is likely lying because honest ads incur no added cost. The 2001 Nobel Prize in Economics acknowledged the tenets of information economics precisely due to the power of "signals"

to separate truth from lies.[60] Informative signals are potentially expensive actions taken by knowledgeable parties that back up their claims. A product sold with a guarantee, for example, is almost always more reliable than a product sold "as is" or "buyer beware." Good sellers, knowing their claims are true, can offer guarantees that bad sellers, knowing their claims are false, cannot afford to offer. The voluntary signal separates good from bad, and fact from fiction. The proposed mechanism is powerful.

This externality signaling market mechanism exhibits several important properties. (i) Knowing a claim to be false, an author will not want to guarantee its veracity. (ii) True claims are costless to the author, enabling honest authors to voluntarily signal by offering a guarantee. (iii) Together, these properties yield a separating equilibrium based on authors' private knowledge, distinguishing misinformation from authoritative information. (iv) Initial burden for deciding truth rests with the author rather than the platform or the uninformed reader. This is more socially efficient as it solves a negative externality problem. Specifically, the decision to warrant / not warrant a claim places initial burden of proof on the polluter rather than on third parties compensating for that pollution, thus it should reduce pollution. (v) Crowdsource detection of falsehoods scales. Members of the crowd are motivated to detect false claims by the reward. (vi) The challenge fee discourages false challenges. It also covers the costs of adjudicating a challenge so the mechanism is financially self-sustaining. (vii) Jury adjudication makes it harder for ideologues to discredit fact checking relative to standing bodies whose verdicts, though true, they dislike. (viii) The entire mechanism is decentralized and market based. This solves the conflict-of-interest agency problem of having the platform or an authoritarian government adjudicate truth.

Interestingly, establishment of a "market for truth" is also business model compatible. It removes responsibility for adjudicating truth from the social media platform, returning this to society, yet it enables the platform to participate in the advertising and escrow markets. Conditional on building the institutional infrastructure necessary to support it, an extended Coasian market-based solution combined with information economics is economically supportable. Such a market, analogous to that for carbon trading, could address the dysfunctional information problem.

Objections

The complexity of intervening in news streams means that inevitably some constraint might be broken. This section seeks to address the most common objections to interventions in any market for free speech. These four are among the most common.

1) *Platforms do not produce the content they propagate.*

[60] Spence, M. (1978). Job market signaling. In *Uncertainty in economics* (pp. 281-306). Academic Press. Akerlof, G. A. (1978). The market for "lemons": Quality uncertainty and the market mechanism. In *Uncertainty in economics* (pp. 235-251). Academic Press. Stiglitz, J. E., & Weiss, A. (1981). Credit rationing in markets with imperfect information. *The American economic review*, *71*(3), 393-410.

If social media platforms do not author the content that cause harm, why should having them internalize these costs be more efficient than having content creators bear these costs? Section 230 grants them immunity on this basis. A more robust analysis provides two answers based on governance and transaction costs. First, social media platforms already internalize the positive externalities of social networks; they need only internalize the negative externalities as well. Social media platforms are built on and derive their power from network effects. Their purpose is to foster connections.[61] The contacts and activity of one user benefit other users. These are externalities and when they are positive the platform already encourages them and profits from them by interposing itself and monetizing ads between connections. When harms occur on-platform, as in the case of harassment or fraudulent products, the platform already addresses them. Platforms only need motivation to take actions they already take yet must do so for harms that occur off-platform in addition to those that occur on-platform.

Although social media platforms do not author the content they dispense, any claim they exert no influence over members is disingenuous. They actively engage in orchestration. They are the governments of their ecosystems with authority to regulate participation, prices, competition, and intellectual property within their regimes.[62] The venture capitalists who invest in platforms, not merely the economists, have stated as much.[63] When the citizens of one country suffer the pollution of another, the government of the former might reasonably negotiate with the government of the latter, especially when polluters in the latter are invisible to citizens of the former. Although neither government itself produced the pollution, lax rules in the source country are at least partially responsible for pollution in the harmed country. In this case, the costs of harmed individuals bargaining with each polluting firm greatly exceed those of collective bargaining. This argues for negotiating with the government of the polluting country as the highest leverage point of intervention.

Second, platforms do, in fact, represent the nexus of lowest transaction costs.[64] Social media platforms, unlike one-way broadcast media, orchestrate the activities of their users. In order to facilitate membership and engagement, they reduce friction on participation and production. They provide tools for creation, tools for consumption, and feedback on impact. In fact, no party has greater visibility than social media platforms into the nature of misinformation transactions. Without information supplied by the platforms, not even the authors themselves know who has shared or who has read their campaigns. Transaction cost economics weigh in favor of intervening at the point

[61] Facebook's original mission statement was "Making the world more open and connected." Constine, J. "Facebook changes mission statement to 'Bring the world closer together'" TechCrunch https://techcrunch.com/2017/06/22/bring-the-world-closer-together/

[62] Parker, G. G., Van Alstyne, M. W., & Choudary, S. P. (2016). *Platform revolution*. Norton. Chapter 8 - Governance. Tirole, J. (2017). *Economics for the common good*. Princeton University Press.

[63] Brad Burnham, "Web Services as Governments," Union Square Ventures, June 10, 2010, https://www.usv.com/blog/web-services-as-governments.

[64] Juliet Yu, Alibaba. Munger, M. (2015). "Coase and the Sharing Economy," in Forever contemporary: the economics of Ronald Coase. Veljanovski, Cento (ed). Institute of Economic Affairs, pp 187-208.

of lowest cost and greatest transparency, in this case, the point of the platform. Social media platforms represent the point of greatest leverage.

2) How can this or any mechanism decide what is really true?

The effort to ascertain truth has two approaches, one practical and one theoretical. In practice, courts routinely grapple with the question of whether a claim is true "beyond a reasonable doubt." They address this challenge by raising the level of threshold to the import of the problem. For example, the legal tests that we apply to free speech laws and their breach must pass three different thresholds. Roughly stated, these are *rational basis review*: does the law relate to a legitimate end and has it in truth been violated? This dispenses with specious cases. A higher level is *intermediate scrutiny:* does violating the law affect a basic right and has it in truth been violated? This protects individual rights. The highest level is *strict scrutiny*: intervention must protect a compelling government interest and be narrowly tailored. This protects group rights. Similarly, the laws around "duty of care" for product liability are particularly vague and differ by state in the same manner that free speech laws differ by country. And yet, as a practical matter, we deal with them. We apply local context and change the threshold for certainty according to the severity of the decision.

As a theoretical matter, we cannot know absolute truth. This objection simply re-asks Hilbert's "*Entscheidungsproblem*" in a new context. The Church-Turing thesis tells us that certain statements cannot be proven true or false. A more precise statement of Hilbert's decision problem, grounded in logic and philosophy, is as follows. Given a system of claims, is it possible to definitively prove the collected assertions are true? The answer, in general, is no. Posed in 1928, this hard question was not answered until 1936, when Alonso Church and Alan Turing independently developed methods to prove that an infinity of claims are undecidable. Modern computer theory provides an interesting clarification. If statements are made at one end of a communication channel, can one be certain that identical statements are received at the other end of that channel? In effect, data corruption – literally false news introduced into the channel – can be repaired using error correcting methods but only up to a point. Shannon's Channel Coding Theorem, which forms the basis of all modern communications, proves that arbitrarily small error in communicating a fact is not achievable. Error correction is only possible up to a fixed and finite boundary. It is impossible past that boundary.

Thus, the question presupposes an answer that cannot be given. This objection is used to dispatch any approach that cannot solve the problem, which is an unfair critique because *no* approach can solve the problem. We can only know truth to a given number of bits. Interestingly, this comports well with the practical solution of accepting a claim as true "beyond a reasonable doubt."

3) *Edge cases between true and false invalidate the mechanism. Whatever the boundary condition, it is always possible to split the boundary with careful wording.*

The existence of an edge case is not a legitimate challenge to any governance mechanism generally, let alone a fair news mechanism specifically. There does not exist any useful mechanism for which there do not exist edge cases between true and false. The decision criterion "Always Guilty" has no edge cases but also no practical application, as does its opposite "Always Innocent". Even our most cherished and most absolute rights have edge cases. Is the right of free speech absolute? We admit slander, libel, and incitement to violence as exceptions. Is the right to life absolute? We admit self-defense as an exception. The existence of edge cases can be used to exclude every mechanism, which leaves only the null mechanism, thus it fails as a legitimate test. Rather, the test should be whether one mechanism adjudicates edge cases better than the alternative. Importantly, the alternative is not the null set of no mechanism at all. In our case, the alternative is the present mechanism being used by social media platforms and, judged in terms of efficacy and absence of bias, that leaves much to be desired. The proper challenge is therefore to articulate the alternative mechanism and show why it does better than the proposal under consideration. Across a weighted sum of false positives and false negatives, which rule achieves more social value? Admittedly, this is a high bar. The best challenge is a superior mechanism design. The best and most challenging objection is thus an act of creation and not merely an act of rejection.

4) *Reasonable people will disagree and those who dislike a decision will simply seek to discredit the decision maker. Unless adjudication is indisputably impartial, partisans will not accept results.*[65] *We cannot avoid the problem of who gets to decide.*

Conservatives may reject a decision whose outcome favors a liberal view. Liberals may reject a verdict whose outcome favors a conservative view. This objection raises separate issues of reconciling opposing views and of decision legitimacy. On the issue of reconciling conflicting views, there are three reasons why *requiring* agreement is ill advised.

1. People do not universally wish to be convinced nor do they grant third parties the moral authority to convince them. They often reject data that disagrees with their identity or world view or position in life. "It is difficult to get a man to understand something when his salary depends on him not understanding it."[66] More carefully, one may reasonably ask, what gives the mechanism designer the moral authority to assert the righteousness or truthfulness of the mechanism's verdict?[67] Absent such authority, perhaps the empowered view should shift its position to the disempowered view.
2. Mechanisms that *require* agreement cause moral hazard. If partisans know they will be bought out, with resources needed to convince them, they can exaggerate their

[65] This objection arose in a conversation with the misinformation team at Facebook.

[66] Sinclair, U. (1994). *I, Candidate for Governor*. University of California Press.

[67] Mill, J. S. (1966). On liberty. In *A selection of his works* (pp. 1-147). Palgrave, London.

claimed protests and supposed beliefs. These beliefs are not themselves verifiable. The social cost to overcome this mechanism-induced moral hazard could be in excess of the value of the verdict, producing social waste. The alternative, coercion, risks reaching the wrong conclusion simply by placing the power of coercion with one or another party.

3. The most compelling reason, however, that requiring agreement is not a valid test is an artifact, again, of the *Entscheidungsproblem*: knowing absolute truth is absolutely impossible. If one *unbiased* party cannot know or even communicate certain truth, it is pointless to require multiple *biased* parties to agree on certain truth. Universal agreement is an impossible standard.

If reconciliation is infeasible, then in what sense might a verdict be legitimate? The solution is one we recognize in other contexts as procedural fairness.[68] Partisans must agree *ex ante* to the method for deciding what's true, then commit to abide by the impartially administered verdict.

Thus, to operate a market for truth, we can rely on established administrative practices that we already use for trust and legitimacy. Taking our own government as precedent, consider a design where we split fake news oversight into legislative, judicial, and executive offices. A legislative body gets to define "fake ads." Despite their differences, even Fox News, CNN, and the New York Times might be able to agree on a working definition of fake news independent of specific cases. A judicial body gets to decide whether a specific case represents an instance of fake news according to this definition. Fact-checking organizations or juries of peers might play this role only now they must judge according to the definition provided by the legislative body. Jurors do not get to use their own individual definitions. Finally, the executive branch enforces these definitions and decisions. Social media platforms like Facebook and Twitter can play this role but they decide neither the definitions nor the outcomes of challenges. By dividing the branches of fake news governance, we recreate an institution where no branch judges truth as applied to itself and no branch has an economic incentive to bias its behavior to get rich. The divided process should therefore be free of conflict of interest, less biased due to random sampling, and by design more legitimate.

Conclusions

This is a work-in-progress distributed to solicit feedback. Pls send comments, support, objections and critiques to mva@bu.edu. Thank you.

[68] Van den Bos, K., Wilke, H. A., & Lind, E. A. (1998). When do we need procedural fairness? The role of trust in authority. *Journal of Personality and social Psychology*, *75*(6), 1449.
[69] National Federation of Independent Businesses et. al. v Sebelius, Secretary of Health and Human Services et. al.