

Folk Models of Misinformation on Social Media

Filipo Sharevski^{1,2}, Amy Devine, Emma Pieroni, Peter Jachim

¹DePaul University, Chicago, IL

²Adversarial Cybersecurity Automation Lab (ACAL)

Network and Distributed Systems (NDSS) Symposium 2023



Models of Misinformation: Fake, Mistake, Calculate?

Fake → News, falsehoods, fabrications, rumors, conspiracies, hoaxes, trolling, memes, clickbait, spam... [29, 30];

Mistake → Inadvertent reporting, calculated omission, incorrect understanding, wrong assumptions.. [23, 19]

Calculate → Disinformation, network propaganda, satire... [1, 27]

Everyone is entitled to their own opinion, but ~~not~~ and to their own facts?



Responses to Misinformation: Humans

Predisposition → Lack of reasoning, ignorance, inattention, favoritism [15, 7]

Perception → Illusory truth effect, implied truth effect, backfire effect [13, 6]

Polarization → Partisanship, ideology, values/preferences/beliefs [16, 3, 17]

Promulgation → Lax platform policies, diminished role of experts [19, 2]

Panic and Pandemonium → moral, interpersonal, technical, societal...



Responses to Misinformation: Scientists and Systems

Prebunking → Forewarning, preemptive refutation, accuracy nudges [12, 14]

Debunking → Verifiable corrections from credible sources [5, 6]

Detecting → Content, metadata, context, engagement [8, 21, 9]

Moderating → Covers, warnings, account removal/suspension [28, 11, 18]

Reacting → *Alt-platforms, forums, messaging boards...*



Models of Misinformation: What's Missing?

What we know → How people **respond** to misinformation

What we see → How platforms **handle** misinformation

What we are curious about → How people **model** misinformation?



Folk Models of Security

Folk models → Not necessarily accurate mental representations, but shared among similar members of a culture [24]

Low comprehension → Mental models for security-related decision

Aspects	Models
Camp: analogies [4]	<i>physical, criminal, medical, warfare and market</i>
Wash: software [25]	<i>bad or buggy software by mischiefs or criminals</i>
Wash: hackers [25]	<i>burglars, bigfish attackers, vandals, contractors</i>
Wash and Rader: ecosystem [26]	A malicious software that <i>originates</i> on Internet, <i>creates</i> visible problems, people could <i>self-protect</i> , <i>everyone</i> is a target



Folk Models of Social Media

Folk models → Descriptions, explanations and predictions of platforms' purpose and participation [24]

Participation → Self-presentation, content historicity, online persona(s)

Aspects	Models
Johnson et al.: <i>exposure</i> [10]	concerned with inappropriately sharing content with members of the friend network
Schaik et al.: <i>engagement</i> [22]	less concerned with password breach, more concerned about who can tag, see and comment their posts
Sleeper et al.: <i>rollback</i> [20]	concerned if revealing too much, direct criticism, direct attack, or blunder and try to repair their stance



Intersection of Folk Models: Misinformation?

Is misinformation, in folk terms,...

Analogies → ... a **threat** for the physical/cybersphere well being?

Software →... a **bad** or **buggy** “platform-ware”?

Hackers →... shared by **vandals, mischiefs** or **(war) criminals**?

Ecosystem → ... something that does **originates** on Internet, **creates** visible problems, people could **self-protect**, **everyone** is a target?

Exposure → ...an **inappropriately shared content** with members of a friend network

Face →.. .about **tagging, seeing**, and **commenting** on profiles?

Rollback → ... **too revealing, direct criticism/attack**, or **blunder** that needs to be rolledback?



Folk Models of Misinformation: Study

Misinformation → Folk Models, Origins, Purpose, Response, Assessment

Sample → 235 participants, MTurk, 40 minutes, \$6.20, anonymous, social media users

Participants → P136FL60 = No. 136, Female, Left-leaning, Age bracket [51-60]

Gender				
Female	Male	Prefer not to say		
102 (43.4%)	117 (49.78%)	16 (6.82%)		
Age				
[18-30]	[31-40]	[41-50]	[51-60]	[61+]
32 (12.76%)	100 (42.55%)	60 (25.53%)	28 (11.91%)	15 (6.38%)
Political Leanings				
Left	Moderate	Right	Apolitical	
115 (48.93%)	61 (25.95%)	49 (20.85%)	10 (4.25%)	
Misinformation encountered on:				
Facebook	Twitter	Reddit	4chan	Other Platforms
156	131	54	17	6



Folk Models of Misinformation



source



Model I: Political (Counter)Argumentation

Misinformation → Any information that has faithfulness to *selective* facts relative to political and ideological contexts, created and disseminated with agenda-setting or argument-winning intentionality

Notable → “Hammering faulty logic into selection of facts as long as that discredits the ‘other’ side and ultimately wins an argument” [P14FR40].

Quotable → “Political statements that exaggerate the party agenda” [P169FL50]

Origins → The **“other side”** and political “echo chambers”

Notable & Quotable → “Uneducated, bigoted, prevaricating, and shameless hypocrites” [P65FL60], “truly deluded, insistent on being stupid” [P69FL61+], “people who refuse to accept reality” [P79MM61+], and “impulsive, uneducated people” [P96FR60].



Model I: Political (Counter)Argumentation (cont.)

Purpose → *“It incites people to hate political opponents for bogus reasons”* [P21FL60] and *“creates a ‘hive mind’ on social media”* [P44ML30]

Notable → *“Keeps the conservatives occupied, encourages discourse”* [P229FL40]

Quotable → *“Creates the illusion that left-wing positions are popular”* [P68MR40]

Responses → *Ignore as mockery, report if harmful, block if too much*

Notable → *“It is important to be aware of the false information that is spreading to be informed of what some others may be thinking, as being ignorant could lead to another January 6th.”* [P164FM60]

Quotable → *“Not amplifying the other agenda by engaging with it”* [P181FL40]



Model II: Out-of-context Narratives

Misinformation → Any information that has *questionable* faithfulness to known facts due to selection of improbable alternative contexts, created and disseminated with speculative intentionality

Notable → “*Cherry picking events presented out of context in order to support a biased argument*” [P28ML40].

Quotable → Alternatives narratives where “*facts with missing, incomplete, or used in a made-up context*” [P200FM30]

Origins → “*Twisting what’s actually a personal opinion into one’s subjective idea of a fact*” [P200FM30]

Notable & Quotable → “*Misinformation comes, I think often, through selective reading. People want to confirm their narrative, and so they take things out of context, or in limited context. In reality, things are usually more complex. But rather than deal with complexity, simplistic takes that confirm pre-existing narrative biases get read (and shared) more on social media.*” [P188MR40]



Model II: Out-of-context Narratives (cont.)

Purpose → “*Stir the pot*” [P61MM50], “*muddy the waters*” [P4FL40], and “*keep people up in arms*” [P13FR50]

Quotable → “*Indoctrination*” [P36MM40], “*create less trust in institutions*” [P166NL60], “*distract from real issues*” [P33FL40]

Notable → “*Misinformation drives page views; Page views generate money; Follow the money.*” [P176MM50]

Responses → Keen on fact checking and engaging with the “*misinformers*” on social media as their “*pledge to refute false assertions*” [P148FL60]

Notable → Out-of-context narratives “*turn social media into a breeding ground for hatred, racism, misogyny and greed*” [P127FL60]

Quotable → “*Making enemies among friends and acquaintances based on difference in interpretation of hot topics*” [P189FL40]



Model III: Inherently Fallacious Information

Misinformation → Any information unfaithful to known facts, regardless of contexts or intentionality

Notable → “Hoaxes that circulate on social media” [P91ML60] that include “*fear mongering information*” [P169FL50]

Quotable → “*Wild ideas*” [P184MM40], “*lies*” [P20FL61+], or “*blatant falsehoods*” [P157FI50]

Origins → spread of falsehoods “*comes from ignorance, hate, and anger*” [P41FL40]

Notable & Quotable → Spreaders of misinformation: “*idiots with subpar IQ*” [P184MM40], “*ignorant people*” [P134FL61+], “*disingenuous people*” [P57FL50], or “*opportunists*” [P38ML60]



Model III: Inherently Fallacious Information (cont.)

Purpose → *“Usually to stir up controversy or sow dissension amongst the masses”*
[P90FL60]

Quotable → *“Brings some sort of anarchy or civil disobedience about so as to hurt people”* [P149MR40]

Notable → *“To try to get people to click on links”* [P141FM50], *“roil them up, and increasing engagement with the platform”* [P140FL60]

Responses → something that can be thwarted by direct action on social media platforms

Notable → *“Always report posts that contain false information, spread hatred or disenchantment; and always block the user as well* [P57FL50]

Quotable → *“Debating only adds credence to something that deserves no attention”*
[P99ML61+]



Model IV: External Propaganda

Misinformation → Any information that *fluctuates* its faithfulness to known facts relative to shifting contexts or perceived division-creating intentions

Notable → A distinct “*propaganda*” [P201FL61+] flavor to the information operations

Quotable → “*Organized propaganda campaigns, meant to undermine the United States*” [P128ML50]

Origins → “*Nation states hostile to the United States and her interests*” [P60MR40]

Misinformation Chain → a) “*The fake accounts controlled by the state actors put initial rumors and fabricated facts*” on social media [P97FR60] → b) these are picked up and amplified by “*demagogue figureheads that glom onto misinformation that suits their needs*” [P60MR40]; → kept alive by “*ignorant individuals freely sharing it*” [P128ML50] that appropriated the misinformation as the “*preferred truth*” [P105ML40]



Model IV: External Propaganda (cont.)

Purpose → “Serving the function of persuading the more naïve into being fear-mongered into extreme beliefs.” [P80FL40]

Quotable → “Destroy liberal democracy, rule of law, and replace it with a kleptocratic form of government where the financial elite no longer have pesky regulations or taxes, minorities have no rights, and the enforcement of law becomes arbitrary” [P60MR40].

Notable → Exploit “impulsive people” [P96FR60] and “idiots” [P112FL40] to “advance their agenda, for example, erode the trust in institutions” [P183ML40]

Responses → “Not to take emotion-provoking posts at the face value” [P80FL40]

Notable → “Overshadow it with information with maximum faithfulness to known facts” [P225FL30]

Quotable → Debunking Russian propaganda in particular, “from Pizzagate to denazification of Ukraine” [P96FR60].



Model V: Entertainment

Misinformation → Any information with a *tangential* faithfulness to known facts relative to humorous or sarcastic contexts, usually created and disseminated with intention to “*entertain*”

Notable → “*Making fun of the misinformation itself*” [P233FM40]

Quotable → Not to (counter)argue with posts perceived as polarizing, but to “*mock off-the-wall posts*” [P123FR50] themselves

Origins → Jokes and memes are used to “*spread a negative view towards a figure, issue, or movement*” [P58FL50]

Social Media Culpability → a) Enable memes to “*spread like wildfire*” [P2FL40] → causing misinformation to come to attention to “*people who are naturally drawn to posts with wild and far-fetched ideas*” [P43MM30]



Model V: Entertainment (cont.)

Purpose → “Doesn’t have a true function besides satire and entertainment” [P78FL40]

Notable & Quotable → “It is fun to watch people describe daily all sorts of impossible stuff that is bothering them.” [P130ML50]

Responses → “It’s not worth arguing with them, only laugh at them” [P216MR50]

Notable → Block, mute, report, comment → “these actions are warranted when memes stop being taken humorously and become seeds for more misinformation” [P111FL50]

Quotable → “Mock how stupid the misinformation is and laugh about it openly” [P125ML40]



Folk Models of Misinformation



source



Immunity to Misinformation

Rule-of-Thumb → *“If it seems insane that is the first clue. Beyond that, I look at the site or the source being used. If there is no source? 90% chance it’s a lie. If there is a source/site listed it doesn’t take much effort to glance at it and know if it’s misinformation or an extremist site”* [P8MA40]

Notable & Quotable → *“Confident to find the truth after years of fine-tuned bullshit detection”* on social media [P135ML40] → fact-checking + scientific evidence

“Shoe on the other foot” → *“Reply with something to make the misinformer think about what they’re doing (e.g. what if this is your mom) and the repercussions they might bring to someone”* [P90FM60].

Notable & Quotable → *“Yes, I blocked my own husband on Facebook because he was spreading misinformation regarding the 2020 election; I will block anyone who does this”* [P53FL40]



Evolution of the Folk Models

Physical and Medical Model → “People have died because of the politicization of a pandemic with misinformation” [P65FL60]

Criminal Model → “Inciting unrest and insurrections in the country” [P20FM30]

Warfare Model → “The Russians for waging psychological warfare against the west for years” [P20FL61+]

Market Model → “Weakening the US economy and military power without fighting a real battle” [P55ML40]

Bad/Buggy “Software” → “Blatant lies” [P71ML60], “speculations” [P123FR50], “fabrications or distortions” [P63MM40], “flawed opinions or logic” [P64MM40]



Evolution of the Folk Models

Burglars → “Entities interested in chaos and money” [P163ML40]

Vandals → “People trying to stir up trouble, spread their lies, looking for attention, or looking to get ‘likes’” [P19FM40]

Contractors/opportunists → “Troll farms and Russian bots” [P32ML50]

Exposure → “A few family members I’ve had to un-follow because they post such ridiculous things,” [P53FL40]

Engagement → “counter misinformation with as much sourced, unbiased information as possible” [P32ML50]

Rollback → “you lose someone [sic] well-intentioned misinformation spreader if you mock or taunt with personal attacks” [P212MR40]



Misinformation Assessment

Analytical Reasoning → *“In a political discussion, especially an argumentative one involving more than one perspective, it’s likely that at least some misinformation is being spread, and this is often self-evident when one side makes one claim and another makes the opposite, as both claim and counter claim can’t be true at the same time.”* [P15FL50]

Precaution → *“Statistics don’t lie but liars use statistics”* [P212MR40]

Tip-off Cues → *“External boogeyman accounts of the Russians containing spelling mistakes, bad grammar, and weird patriotic sounding name”* [P230MR50]

Tip-off Cues → *“Bunch of numbers, generic names, and fake images in the profile, something like Tom87654.”* [P162ML50]

Gut Feelings → *“If something gets an emotional reaction out of you, it is time to question the veracity.”* [P86ML50]



Thank you!

Questions, Comments, Concerns



Twitter: @ACALaboratory

References I

- [1] Yochai Benkler, Robert Faris, and Hal Roberts. *Network propaganda: Manipulation, disinformation, and radicalization in American politics*. Oxford University Press, 2018.
- [2] Hal Berghel. “Malice domestic: The Cambridge analytica dystopia”. In: *Computer* 51.5 (2018), pp. 84–89.
- [3] Shweta Bhatt et al. “Illuminating an Ecosystem of Partisan Websites”. In: *Companion Proceedings of the The Web Conference 2018*. WWW ’18. Lyon, France: International World Wide Web Conferences Steering Committee, 2018, pp. 545–554. ISBN: 9781450356404. DOI: 10.1145/3184558.3188725. URL: <https://doi.org/10.1145/3184558.3188725>.
- [4] L. Jean Camp. “Mental models of privacy and security”. In: *IEEE Technology and Society Magazine* 28.3 (2009), pp. 37–46.
- [5] Man-pui Sally Chan et al. “Debunking: A Meta-Analysis of the Psychological Efficacy of Messages Countering Misinformation”. In: *Psychological Science* 28.11 (2017), pp. 1531–1546.
- [6] Ullrich K. H. Ecker et al. “The psychological drivers of misinformation belief and its resistance to correction”. In: *Nature Reviews Psychology* 1.1 (2022), pp. 13–29.
- [7] Jr Edson C Tandoc, Darren Lim, and Rich Ling. “Diffusion of disinformation: How social media users respond to fake news and why”. In: *Journalism* 21.3 (2020), pp. 381–398. DOI: 10.1177/1464884919868325. eprint: <https://doi.org/10.1177/1464884919868325>. URL: <https://doi.org/10.1177/1464884919868325>.



References II

- [8] Austin Hounsel et al. “Identifying Disinformation Websites Using Infrastructure Features”. In: *10th USENIX Workshop on Free and Open Communications on the Internet (FOCI 20)*. USENIX Association, Aug. 2020. URL: <https://www.usenix.org/conference/foci20/presentation/hounsel>.
- [9] Peter Jachim, Filip Sharevski, and Paige Treebridge. “TrollHunter [Evader]: Automated Detection [Evasion] of Twitter Trolls During the COVID-19 Pandemic”. In: *New Security Paradigms Workshop 2020*. NSPW '20. New York, NY, USA: Association for Computing Machinery, 2020, pp. 59–75.
- [10] Maritza Johnson, Serge Egelman, and Steven M. Bellovin. “Facebook and Privacy: It’s Complicated”. In: *Proceedings of the Eighth Symposium on Usable Privacy and Security*. SOUPS '12. Washington, D.C.: Association for Computing Machinery, 2012. ISBN: 9781450315326. DOI: 10.1145/2335356.2335369. URL: <https://doi.org/10.1145/2335356.2335369>.
- [11] Jan Kirchner and Christian Reuter. “Countering Fake News: A Comparison of Possible Solutions Regarding User Acceptance and Effectiveness”. In: *Proc. ACM Hum.-Comput. Interact.* 4.CSCW2 (Oct. 2020).
- [12] Stephan Lewandowsky and Sander van der Linden. “Countering Misinformation and Fake News Through Inoculation and Prebunking”. In: *European Review of Social Psychology* 32.2 (2021), pp. 348–384.



Stephan Lewandowsky et al. *The Debunking Handbook 2020*. 2020.

References III

- [14] Rakoen Maertens et al. “Long-term effectiveness of inoculation against misinformation: Three longitudinal experiments.”. In: *Journal of Experimental Psychology: Applied* 27.1 (2021), p. 1.
- [15] Gordon Pennycook and David G. Rand. “The Psychology of Fake News”. In: *Trends in Cognitive Sciences* 25.5 (2021), pp. 388–402.
- [16] Martin Potthast et al. “A stylometric inquiry into hyperpartisan and fake news”. In: *arXiv preprint arXiv:1702.05638* (2017).
- [17] Filipo Sharevski et al. “Gettr-ing” User Insights from the Social Network Gettr.
https://truthandtrustonline.com/wp-content/uploads/2022/10/TT0_2022_proceedings.pdf. Boston, MA, 2022.
- [18] Filipo Sharevski et al. “Meaningful Context, a Red Flag, or Both? Preferences for Enhanced Misinformation Warnings Among US Twitter Users”. In: *Proceedings of the 2022 European Symposium on Usable Security*. EuroUSEC ’22.
<https://doi.org/10.1145/3549015.3555671>. Karlsruhe, Germany: Association for Computing Machinery, 2022, pp. 189–201. ISBN: 9781450397001. DOI: 10.1145/3549015.3555671.
- [19] Filipo Sharevski et al. “VoxPop: An Experimental Social Media Platform for Calibrated (Mis)Information Discourse”. In: *New Security Paradigms Workshop*. NSPW ’21. Virtual Event, USA: Association for Computing Machinery, 2021, pp. 88–107. ISBN: 9781450385732. DOI: 10.1145/3498891.3498893. URL: <https://doi.org/10.1145/3498891.3498893>.



References IV

- [20] Manya Sleeper et al. ““I Read My Twitter the next Morning and Was Astonished”: A Conversational Perspective on Twitter Regrets”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '13. Paris, France: Association for Computing Machinery, 2013, pp. 3277–3286. ISBN: 9781450318990. DOI: 10.1145/2470654.2466448. URL: <https://doi.org/10.1145/2470654.2466448>.
- [21] Gianluca Stringhini. “Computational Methods to Understand and Mitigate Online Aggression”. In: USENIX Association, Feb. 2021.
- [22] Paul van Schaik et al. “Security and privacy in online social networking: Risk perceptions and precautionary behaviour”. In: *Computers in Human Behavior* 78 (2018), pp. 283–297.
- [23] Luis Vargas, Patrick Emami, and Patrick Traynor. “On the Detection of Disinformation Campaign Activity with Network Analysis”. In: *Proceedings of the 2020 ACM SIGSAC Conference on Cloud Computing Security Workshop*. CCSW'20. Virtual Event, USA: Association for Computing Machinery, 2020, pp. 133–146. ISBN: 9781450380843. DOI: 10.1145/3411495.3421363. URL: <https://doi.org/10.1145/3411495.3421363>.
- [24] Melanie Volkamer and Karen Renaud. “Mental Models – General Introduction and Review of Their Application to Human-Centred Security”. In: ed. by Marc Fischlin and Stefan Katzenbeisser. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 255–280.



References V

- [25] Rick Wash. “Folk Models of Home Computer Security”. In: *Proceedings of the Sixth Symposium on Usable Privacy and Security*. SOUPS '10. Redmond, Washington, USA: Association for Computing Machinery, 2010. ISBN: 9781450302647. DOI: 10.1145/1837110.1837125. URL: <https://doi.org/10.1145/1837110.1837125>.
- [26] Rick Wash and Emilee Rader. “Too Much Knowledge? Security Beliefs and Protective Behaviors Among United States Internet Users”. In: *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*. Ottawa: USENIX Association, July 2015, pp. 309–325. ISBN: 978-1-931971-249. URL: <https://www.usenix.org/conference/soups2015/proceedings/presentation/wash>.
- [27] Tom Wilson, Kaitlyn Zhou, and Kate Starbird. “Assembling Strategic Narratives: Information Operations as Collaborative Work within an Online Community”. In: *Proc. ACM Hum.-Comput. Interact.* 2.CSCW (Nov. 2018).
- [28] Thomas Wood and Ethan Porter. “The Elusive Backfire Effect: Mass Attitudes’ Steadfast Factual Adherence”. In: *Political Behavior* 41.1 (2019), pp. 135–163. DOI: 10.1007/s11109-018-9443-y. URL: <https://doi.org/10.1007/s11109-018-9443-y>.
- [29] Liang Wu et al. “Misinformation in Social Media: Definition, Manipulation, and Detection”. In: *SIGKDD Explor. Newsl.* 21.2 (Nov. 2019), pp. 80–90. ISSN: 1931-0145. DOI: 10.1145/3373464.3373475. URL: <https://doi.org/10.1145/3373464.3373475>.



- [30] Savvas Zannettou et al. “The Web of False Information: Rumors, Fake News, Hoaxes, Clickbait, and Various Other Shenanigans”. In: *J. Data and Information Quality* 11.3 (May 2019). ISSN: 1936-1955. DOI: 10.1145/3309699. URL: <https://doi.org/10.1145/3309699>.

