[Pre-publication draft: final version forthcoming in Popa-Wyatt, ed, *Misinformation and other Epistemic Pathologies*, Cambridge University Press.]

Signal-to-Noise: Toward an Alternative Account of the 'Fake News' Phenomenon

Endre Begby, Simon Fraser University

Abstract: Typical discussions in philosophy and elsewhere work on the assumption that Fake News targets our democratic processes by way of exploiting known weaknesses in our individual epistemic defenses: taking aim at our naïve and gullible patterns of news consumption, it aims to trick us into adopting new false beliefs, or, minimally, to become less confident in the beliefs that we already have. In short, Fake News is designed to persuade, or else, to confuse, to sow doubt, or instill skepticism.

This paper explores a different approach to the phenomenon: Fake News can have a detrimental impact on our democratic processes without persuading or confusing anyone, but simply by driving down the signal-to-noise ratio in our information economy. It can thereby contribute to wide-scale disengagement from the most inclusive platforms of information-sharing in favor of more selective engagement on platforms where one can expect a higher signal-to-noise ratio. As a result, our public deliberations become less inclusive, less representative, plausibly fanning processes of belief polarization, with predictably negative effects on our baseline democratic processes.

I argue that this account fits better empirically with much of what we know about the Fake News phenomenon and also draws attention to issues that typically fall under the radar of discussion. In particular, it calls into the question the discourse centered on creating public awareness of a 'Fake News epidemic': by encouraging us to think of our fellow citizens as potential "disease vectors" in the spreading of disinformation, such rhetoric entrenches the idea that we would do better, epistemically speaking, by significantly shrinking our 'epistemic bubble' – i.e., the set of people in whom we regularly place our epistemic trust – to the point of its becoming a reinforced 'echo chamber.'

1. Introduction

Public concern about Fake News (FN) seems to wax and wane with high-profile political events like wars and elections. Nonetheless, the perception that FN poses a continuing threat to our society seems to have ingrained itself on our public psyche: for instance, a recent World Economic Forum report (2024) lists 'Misinformation and Disinformation' as the most severe global risk we are facing in a 2-year perspective and as the fifth-most severe in a 10-year perspective. To put this in context, this means respondents rank the global dangers of disinformation on a scale with those of climate change (and associated extreme weather events), inter-state wars, and economic insecurity.

Now, then, seems like a good time to ask ourselves how well we really understood this phenomenon in the first place. Standard analyses affirm an attractively clear and tidy answer: Fake News is a politically motivated campaign to undermine our democratic institutions by way of undermining that shared trust in a common body of information that democratic deliberation depends on.[1] It works by exploiting democratic citizens' naïve and gullible patterns of news consumption. In brief, human beings are naturally trusting. FN presents itself as true and travels along communication channels that we normally take to be trustworthy. And so, despite the often outrageous content, we sometimes fall for the ruse and mistake the FN for the real thing. We thereby end up with false beliefs, and false beliefs are presumably bad for deliberation and decision-making, whether at the individual or collective level.

In this sense, standard discussions of Fake News, whether in academic philosophy or in public discourse at large, tend to focus on FN's potential to instill false beliefs, or, more broadly, to create confusion, sow doubt, or generate skepticism. And if that's the way we think about it, then it's also quite reasonable to think that the best way to deal with FN is by deploying a two-pronged strategy: (i) to seek institutional changes aimed at curbing the spread of FN; (ii) to develop awareness campaigns aimed at improving people's 'cognitive resistance' or 'immunity' to FN, so that they are less likely to be fooled.

If there's any glimmer of hope to detect in these discussions, it's because we think this combination of remedies may have actually done us some good: we are now both (i) less likely to be exposed to FN and (ii) individually better / smarter / wiser in how we approach what FN might still manage to sneak through the institutional barriers. If future FN crises can be averted (or ameliorated), the solution will, again, be more of the same: curb the spread and increase immunity.

Maybe there's some truth in this way of approaching FN. But I think there's also room to explore an alternative account. This account would probe more deeply into FN's role with respect to the epistemology of democracy in particular, and would argue that FN could have a significant and detrimental impact on the quality of our democratic processes even if it weren't designed to deceive or confuse anyone in the way that standard discussions presuppose.

Here's what I have in mind: the quality of our democracy depends on the quality of our public deliberations. In turn, the quality of our public deliberations depends on the quality of our shared information environment. By pumping large amounts of cheaply produced FN into our information environment, politically subversive agents can succeed in significantly altering the *signal-to-noise ratio* in our information economy. In doing so, they also significantly reduce the expected utility of engaging in public deliberation on the broadest and most inclusive platforms, and instead motivate us to turn toward

[1] Cf. Ecker et al. 2024.

more selective platforms, where we can expect a higher signal-to-noise ratio. This all but ensures that we will be pushed in a direction where we are more likely to interact with people who already largely share our cognitive outlooks and our sense of value.

In this way, even low-quality FN, produced cheaply on relevant scale, can be a significant contributor to the process of *belief polarization*, which has a well-documented detrimental impact on the quality of democratic lives.[2] As I will argue, it is really the quantity and not the quality that does the trick here. High-quality FN, plausibly designed to deceive even fairly rational and astute news-consumers may well exist. The problem is that it is hard to produce on a scale that might have an impact on the outcome of our public deliberations. By contrast, low-quality FN might have similarly subversive ends, but can achieve those ends at a much lower cost. One observation that sometimes strikes commentators is just *how bad* (as in, how implausible and poorly designed) much of FN seems to be.[3] It's tempting to think, of course, that low-quality FN is just a poor attempt at making high-quality FN. To the contrary, I will argue that low-quality FN carries its own distinctive rationale: it aims not to deceive us or confuse us, but rather to create and exacerbate divisions in our political community.

This account has upshots not just for our understanding of the FN phenomenon itself but can also shed new light on the standard suite of remedial strategies. On the one hand, attempting to curb the spread of FN by fact-checking or credentialing risks de-legitimizing the institution or platform in the eyes of many users. (Especially so, of course, in a socio-epistemic setting that is already significantly marked by belief polarization and the distrust of the 'others' that naturally comes along with it.)

On the other hand, public discourse centered on creating awareness of a 'Fake News epidemic' might well have played its part in causing an ironic backlash: by entrenching the idea that a significant portion of our fellow citizens are potential 'disease vectors' in the spreading of disinformation, it encourages the notion that we might be better off, epistemically speaking, by significantly shrinking our 'epistemic bubble,' perhaps to the point of its becoming a reinforced 'echo chamber.'[4]

When we think about FN in terms of its power to persuade, it is natural to think of it as a matter of individual epistemic failings. But once we move past that perspective, we can see, rather, how the *collective* predicament that our 'FN crisis' has brought about may not be best understood as rooted in failures of individual epistemic rationality at all; instead, it can be seen, in no small part, as driven precisely by

[2] Cf. Sunstein 2002; Talisse 2021; Begby 2022.
[3] See, in particular, Rini 2021, to which I will return later.
[4] For a recent and influential account of these concepts, see Nguyen 2020.

individuals' quite sensible preoccupation with maintaining their own epistemic hygiene in what they reasonably take to be an *antagonistic* (or broadly 'hostile') information environment.

2. Fake news and the epistemology of democracy

'Fake News' can mean many things. For the purposes of this paper, I will assume that our concern is not just with any widely circulated bit of disinformation (such as celebrity gossip). Nor is it just a question of whether it carries *political content*. Instead, I take it that we usually think of FN as organized disinformation campaigns which *target our political agency* in a particular sort of way, thereby aiming to affect the shape and outcome of our public deliberations.[5]

This also helps explain why FN seems to be of particular concern in democratic contexts. Democracies are political regimes that institutionalize mechanisms by which public opinion, via public information sharing, joint deliberation, and voting, directly determines who wields power while also placing constraints on how they wield this power. Because of the high stakes of democratic decision-making, it is therefore not a surprise that many types of agents should take an interest in manipulating that public opinion. And it's natural to think of FN as doing just that, i.e., seeking to cause people either adopt new (false) beliefs or to become less confident in the beliefs that they already have.[6]

Some philosophical background will be helpful in framing my concerns with this picture. I take it that human beings are constitutionally epistemically dependent on others. Considered as individuals, our access to information is very poor: most of what we can know, we can know only by way of our social relations to other people. This epistemic dependence introduces a distinctive kind of vulnerability founded on trust that ought to be at the front and center of our conception of epistemic normativity: it is, quite simply, part of the human predicament.

But even so, not all manifestations of (epistemic) trust are rational: we cannot choose *whether* to trust, but we have a crucial job on our hands in determining *who* to trust, or *how widely* to trust.

Much work in this area permits itself to simplify things a bit, by adopting the idealizing assumption that co-dependent epistemic agents have reason to believe that others are 'acting collaboratively,' i.e.,

[5] To be clear, this is not an attempt to provide a *definition* of FN, like so much of the literature does (cf. Gelfert 2018; Pepp et al. 2019; Anderau 2021; Jaster and Lanius 2021). FN is a rich and complex phenomenon, and I doubt that any definition broad enough to cover all of it would still be philosophically interesting. Instead, this is an attempt to isolate a strand of the FN phenomenon that I believe is particularly relevant to the epistemology of public deliberation and decision-making, and therefore, to the epistemology of democracy.

[6] Whether previous FN campaigns *in fact* had any measurable impact on our public deliberations or election outcomes, by way of any plausible and discernible causal mechanism, is a separate matter: for relevant discussions, see Nelson and Taneja 2018; Guess et al. 2020; Altay and Acerbi 2023.

sharing their best information and generally looking out for each others' epistemic well-being. That is, we take ourselves to have a collective interest in finding out the truth about some subject matter. We communicate honestly, openly, transparently, with a view to sharing all our information, thereby helping to overcome the epistemic limitations that we all suffer as individuals.

This idealization is particularly prevalent in political epistemology. Suppose we were to consider democracy simply on its merits as a collective epistemic decision-procedure: why assume that democracy produces better 'political judgment' than competing electoral systems? A standard answer to this question – call it 'Democratic Optimism' – holds that democracy reliably produces better judgments because in a well-functioning democracy, information-sharing and deliberation among citizens is maximally inclusive, incorporating the widest possible range of epistemic perspectives. So even though the average epistemic competence might be lower in a democracy than it would be in some more selective, 'elite' electoral system, democracy's ability to harness cognitive diversity more than compensates for that deficit.[7]

This is obviously an attractive view, and it's not my task here to argue directly for or against it. Instead, I want to call attention to the underlying idealization: Democratic Optimism basically concedes that in order to vindicate democracy in epistemic terms, we must suppose that democratic citizens are willing and capable of engaging in information sharing and joint deliberation on the scale that the theory supposes. That this *is* an idealization cannot be doubted. For while politics certainly *can* be a venue for open information sharing and joint deliberation toward shared ends, we also know that it often involves the opposite: politics, at scale, is also a battle for power and scarce resources, which incentivizes strategic communication aimed at undermining others' political agency. In recognition of this, much of my own current work – this paper included – is devoted to arguing that we need to develop an approach to epistemic normativity fit to guide us in what I call *antagonistic information environments*.[8]

This still leaves open the question of how FN works: surveying the literature, the overwhelming, and obviously very natural, assumption is that it works by targeting our first-order beliefs and credences. In this sense, it differs from, say, sloppy journalism, but resembles propaganda, with the important difference that whereas propaganda is typically centralized, and relies on the authority of governments or other aspects of the state apparatus, FN is distinctively decentralized, and typically relies on peer-to-peer sharing on social media.

[7] For important examples, see Anderson 2006; Landemore 2012. For a counterpoint, see Brennan 2016.
[8] Cf. Begby 2021*b*; 2022.

To the contrary, I will argue that much of FN is plausibly *not* primarily designed to change our minds in this way.[9] Instead, there are design-features of FN which suggest that it steers by a different goal, roughly, that of causing people to disengage from the political process altogether, or to engage in such processes only in a more selective, less inclusive way. While this is in some sense a less ambitious aim, it is still potentially politically corrosive. More importantly, it is an aim that is much more likely to succeed. In fact, I think it has to some extent *already succeeded*, even as we might congratulate ourselves that we have navigated past our first FN crisis, and that we are now wiser, smarter, and better equipped to withstand the next assault.

3. Fake News: a basic sender-receiver model

Nonetheless, the idea that FN is designed to persuade or to mislead comes so naturally to us that it's hard to think beyond it. After all, what's the point of communication, if not to change people's minds and thereby to nudge their behaviors in particular directions?

To see how we might nonetheless move beyond this assumption, consider your basic sender-receiver model in communication theory.[10] The sender formulates and broadcasts a message. This comes at a certain cost. We want to know why the sender is willing to undertake that cost. Typically, the answer is that they want us to form a certain belief, namely the belief that forms the content of the message.

How about the receiver? First, they need to retrieve and decode the message, secondly, to decide whether to believe it. Most accounts assume that if the receiver is 'tuning in' at all, i.e., listening, it's because they have a disposition to trust the sender. That is, since there's already a cost to listening, we can suppose that they are, at least in typical cases, minimally favorably disposed toward believing what the speaker is about to say.

In a climate marked by a suitable degree of epistemic trust, then, the mere fact that the sender is willing to take on the cost of broadcasting the message already speaks in favor of forming the relevant

[9] It's a different question whether FN-peddlers would be *happy* if they also succeeded in that aim. Much of the literature here is couched in terms of whether FN is 'intended to deceive' (see, e.g., Rini 2017; Dentith 2017; Gelfert 2018; Anderau 2021, and for critical discussion, Pepp et al. 2019). Personally, I don't think 'intended to X' is a sufficiently precise notion for this to be a useful debate: often, we can be said to 'intend' any broadly foreseeable (even if unlikely) beneficial consequences of our actions. For this reason, I prefer to consider the issue in terms of 'design.' (Of course, Pepp et al. are reasonably construed as arguing that neither 'intended to deceive' or 'designed to deceive' should be part of a *definition* of FN. I'm happy to concede the point. But that just underscores how my argument does not proceed from a definition of FN, but rather aims to isolate a particular (and important) strand of the FN-phenomenon.)

[10] Cf. Schramm 1954.

belief. So, in the absence of countervailing considerations, it can be reasonable for the receiver to form that belief merely on the basis of the sender's having broadcast the message.

That's roughly how the standard model works in the context of the epistemology of testimony.[11] Of course, some things are going to be different in the context of media messaging than in face-to-face communication. But the basic model still seems readily applicable to FN: the sender forms a message that they know to be false, but packs it in a manner that resembles the trappings of traditional media communication, which holds a certain kind of authority in our society. The receiver, being naturally trusting, mistakes the FN for the real deal and forms the corresponding belief accordingly.

Even this rather rudimentary outline should be sufficient to highlight the fact that the model under consideration builds in a crucial sort of rationality-assumption: because of the associated costs, we should assume that the message is 'optimized for uptake,' i.e., generally fit to persuade its target audience.

Here's where the model is eventually going to run into trouble. Of course, the story is going to fit *some* FN: in general, FN *does* seem to be disguising its false messages as Real News bulletins, thereby presumably hoping to piggyback on the prestige of institutionalized journalism in our society. And when it succeeds, it seems plausible that it succeeds because it is able to exploit our trust in this institution, and because it has packaged its product in a sufficiently clever way that it might fool even a fairly astute recipient.

This general model seems to be the starting point of familiar 'Public Awareness' campaigns against FN: these campaigns are designed to strengthen our (individual) epistemic defenses against FN by providing an easily implementable checklist for spotting the counterfeit product (e.g., does it come from a major news network?; is it written in good English?, is it credited to named journalists?; is the story confirmed on other news sites?).[12]

By contrast, if FN didn't aim to persuade in the first place, it's not clear what those campaigns are for. So here's the point: I'm not convinced that all or even most of FN really is aimed at persuading in this sense. To be sure, some people will be fooled; conceivably, these kinds of public awareness campaigns might be of help to them. But that doesn't mean that we have 'solved the problem' of FN whenever we have ensured that public 'immunity' has reached a certain threshold.[13]

[11] For a similar perspective from communication theory, see Levine 2022.
[12] Cf. "Ten Questions for Fake News detection."
[13] In fact, there's reason to doubt we were ever below that threshold: for discussion and references, see Altay, et al. 2023.

To see why, consider again the rationality assumption that underlies the model. Empirically speaking, it seems like a *lot* of FN is manifestly bogus and shouldn't be fit to persuade anyone. As Regina Rini points out (2021), it looks more like 'trolling' than an outright attempt at lying or misleading: it is almost like they *want* to be caught in the act.

But that's a bad fit for the rationality-assumptions underlying the standard model: aiming for persuasion requires optimizing your message for plausible uptake on the receiver's end. But so much of FN just doesn't seem optimized in this way at all. Moreover, if the public awareness campaigns were on to something, you would assume that FN-creators would be motivated to improve their product and fix these fairly basic mistakes. Surprisingly, then, much of FN remains manifestly bogus to this day; I'm not aware of any studies that document significant improvement in FN over time.

In short, the rationality assumption underlying the standard sender-receiver model struggles on both ends: it struggles to explain how senders could assume that *this* message, shaped in *this* way, would be fit to persuade any minimally rational receiver. And it struggles on the receiver-end to explain how relevantly large numbers of people could be so gullible as to mistake it for the genuine article.

At this point, I think it will be useful to compare the FN phenomenon with another internet-based scam that has caused a bit of public panic in previous years, the namely the so-called 'advance fee' or 'pay upfront' email scam.[14] The basic design of the scam should be familiar enough: you receive an email claiming to be from some high-end banker or government functionary in a war-torn or highly corrupt country. They tell you that they've stashed away a large amount of money but need help getting it out of the country. They are now offering to share with you the several million-dollar windfall, if only you can first help cover the wire transfer fees: so please deposit some substantial-but-manageable dollar amount to this bank account.

Like FN, the advance-fee email-scam was widely disseminated, and apparently took some people in. Once it was clear that people were falling for it, we then had cycles of public awareness campaigns, not unlike what we saw in response to FN years later. These campaigns might well have succeeded in reducing the absolute number of people who fell for the scam, even if some people still do.[15]

But what's interesting is this: you'd think that in the face of these public awareness campaigns, the scammers would adapt their strategy so as to retain the upper hand. But the basic form of the advance-fee

[14] Also known as the 'Nigerian Prince scam.'
[15] Cf. "'Nigerian prince' email scams still rake in over $700,000 a year—here's how to protect yourself."

scam doesn't seem to have changed all that much: I still receive emails with essentially the same message as I did 20 years ago.

Why haven't they updated their strategy? It seems like it would be simple enough thing to do so, and if you're going to perpetrate a scam, why not try to do it well? Here's my take: while it's crucial that the message be widely distributed, there's a potential bottleneck on the other end that must be kept in mind: any respondents will require substantial amount of extra, individualized handling before they commit to transferring the money. As such, there's capacity limitations involved here: given the amount of extra work required, the scammers would be quickly overwhelmed if a significant portion of the recipients actually responded.

This suggests that the relative transparency of the initial ruse acts as a filter, ensuring that only people with exceptionally low epistemic self-awareness will respond. These are the people that you have the best chance of pulling across the line in the first place. An improved initial 'ruse' would simply be a less effective filter.[16]

In short, the advance-fee scam clearly does *aim at persuasion*: but it can afford to *aim low*, since they only need one person at a time to fall for it, and the less epistemically self-aware they are, the better. Because of superficial similarities, it's tempting to model our understanding of FN on our understanding of the advance-fee scam. But I think it's a bad model, because the baseline aims and expectations that ground sender-rationality are very different in the two cases. As a *financial* scam, the advance-fee scam clearly does aim at persuasion, but not wide persuasion: it can be counted as a success even with very low rates of uptake, which is why we don't see a clear improvement of the product over time. By contrast, FN is a *political* intervention: to have any effect at all, it can't just persuade one person or two persons, but would have to move a critical mass of people. As such, *if* it were aiming to persuade, we should expect to see an improvement to the product over time. But it's not clear that we do, anymore than we do with the advance-fee scam.

If we want to retain a basic sense of sender-rationality, this requires an explanation. And here's of course the explanation that I would like to explore: much of FN really doesn't aim at persuasion in the first place.

[16] Cf. Herley 2012, who argues, from a game-theoretic point of view, that since the scammers have "an over-riding need to reduce false positives," the relative transparency of the initial email is, as it were, a feature, not a bug: "sending an email that repels all but the most gullible the scammer gets the most promising marks to self-select, and tilts the true to false positive ratio in [their] favor."

4. Signal-to-noise ratios in our information economy

But if FN doesn't aim at persuasion, what's the point? My take: FN can have political efficacy by reducing the *signal-to-noise ratio* in our public information economy, thereby making sustained engagement in information sharing and joint deliberation less rewarding, therefore less likely to happen on the scale that we aspire to and that we assume is essential for democratic processes to deliver on their epistemic promise.

Engaging in joint information sharing and public deliberation requires effort. We measure the degree of effort required in part in terms of the cognitive costs of retrieving the signal. The cognitive cost, in turn, is measured with reference to the 'signal-to-noise' ratio.

Think of the 'signal' as the information (or 'knowledge') that one might hope to gain by decoding a message. At all times, there will be a certain amount of 'noise' mixed in with the signal, which needs to be filtered out. While the 'noise floor' can be higher or lower, it would be unrealistic to assume that it could ever be reduced to zero. (That is, I take it that noise is a pervasive element in public epistemology, social or political: some of this noise will come from confusion, misdirection, copying error, idle chatter, or other distractions that invariably intrude on public deliberation. But some of it might also be intentionally placed there precisely to dilute the signal, thereby making it harder to retrieve.[17])

The rationality of engaging in this process is presumably not independent of the underlying cost-benefit analysis, which we can roughly think of in terms of the expected value of retrieving the signal versus the cognitive cost of filtering out the noise. This is politically relevant for the following reason: the quality of our democratic decision-making processes depends on high levels of engagement across a wide range of demographics. For this to happen, we rely on good channels of information sharing and joint deliberation. How well these channels work depends in part on the signal-to-noise ratio. By substantially increasing the noise floor, we make these communication channels less viable as platforms for broad political engagement.

This alternative model can help explain the prevalence of low-quality FN without sacrificing the underlying rationality-assumptions. Simply: creating FN fit to *persuade* a critical mass of potential voters is *hard*. By contrast, it's comparatively easy to pump large quantities of cheaply produced disinformation into our regular channels of political communication. As such, this strategy can plausibly achieve a politically corrosive aim at a much lower cost.[18]

---

[17] The latter, as I conceive of it, is essentially what makes the difference between an 'antagonistic information environment' (cf. Begby 2022) and a merely 'non-ideal information environment' (Begby 2021a).

[18] Don't we at least need to suppose that the FN will be plausible enough to generate a significant amount of engagement? How else would it gain foothold on social media? The problem with this line of questioning lies with our default supposition that such engagement reliably signals belief or credence. On this, see, e.g., Wagner and Boszkowki 2019.

We can imagine a spectrum of outcomes here. One is that people just give up on the whole business of political participation, effectively adopting the cynical attitude that 'all politics is bullshit.' But such self-disenfranchisement is not the only possible outcome. For we can also imagine agents remaining politically engaged but withdrawing from the most inclusive and open-ended fora in favor of narrower, more selective platforms, where they can expect a higher signal-to-noise ratio.

This would entail foregoing the presumptive epistemic benefits of maximally inclusive political engagement, and instead retreating to what Cass Sunstein, in an important early paper on this topic (Sunstein 2002), called 'enclave deliberation': i.e., a pattern where we end up only sharing information and deliberating with people whose opinions already more or less match our own.

As these peer-groups consolidate, we might call the product of this process an 'echo chamber' and we might expect to see patterns of 'belief polarization' emerge over time: the different groups – or 'enclaves' – will become progressively more convinced that their view is right, simply as a consequence of primarily being engaged with views that resemble their own.[19] As a result, the different groups will start to grow apart, each moving in a more extreme direction, jointly leaving behind a yawning gap where the 'moderate opinion' was supposed to be. This is bad news for our collective decision-making because it's presumably the moderate opinion that is more likely to serve as the ground for workable democratic compromises.[20]

In my view, this would be a significant political end. And as such, we shouldn't be surprised to see subversive agents try to pursue it. But notice how it can seemingly be achieved without causing us to adopt any particular new false beliefs or to lose confidence in the beliefs that we already had.

5. Comparisons with extant views

I'm not the first philosopher to suggest that FN might be taking aim at the quality of our information environment, in something like this sense. In fact, the current literature is awash with environmental metaphors such as 'epistemic flooding' (Anderau 2023) or 'epistemic pollution' (Dutilh Novaes and De Ridder 2021). The similarity is plain to see: a 'polluted' epistemic environment is a 'noisy' epistemic environment.

[19] Cf. Begby 2022.

[20] I here purposely omit any discussion of whether the moderate opinion is also more likely to be *true*. I doubt there's any sound generalization across political contexts and times that would vindicate such a supposition. At any rate, nothing in the present argument hangs on this.

Even so, there's a crucial difference in the direction taken by the subsequent analyses. For these accounts will still tend to measure the success of FN in terms its impact on our beliefs or credences. True, instead of simply looking at FN's power to instill new false beliefs, they are now also expanding to consider its ability to 'sow doubt' or to generate confusion and uncertainty. For a poignant example, consider Michael Lynch (2016), writing in the *New York Times*:

> The use of social media to spread political misinformation online is partly just a giant shell game. Propagandists often don't care whether everyone, or even most people, really believe the specific things they are selling (although it turns out that lots of people always do). They don't have to get you to actually believe the penny is under the wrong shell. They just have to get you confused enough so that you don't know what is true.

This is surely an improvement over a simpler model. But it's clear that the account still measures the threat of FN in terms of its direct doxastic effects, i.e., its impact on our beliefs or credences. For instance, reducing the signal-to-noise ratio can make it harder to acquire justified true beliefs, simply because our standard testimonial belief acquisition procedures are rendered *less reliable* by the vast amount of counterfeit product saturating our information environment (cf. Buzzell and Rini 2022). Alternatively, it makes it harder to retain one's epistemic justification because of the large number of (misleading) defeaters that are now percolating about (cf. Blake-Turner 2020). Either way, it's assumed that the degradation of our epistemic environment has a direct and measurable impact on us as individual believers: even if it doesn't cause us to adopt particular false beliefs, it should cause us to lose confidence in whatever beliefs we have.

Perhaps the sharpest analysis here is due to Regina Rini, who claims that the aim of FN is to 'weaponize skepticism.' Rini writes (2021: 37): 'The point was not to gain belief in any one falsehood, but to saturate the epistemic environment with conflicting accounts so that the truth appears to be only one of so many bickering narratives. The goal was confusion and disarray, generalized skepticism rather than focused false belief.'

But closer analysis suggests that there are actually two rather different outcomes that are run together here ('confusion and disarray' versus 'generalized skepticism'), yielding two rather different predictions for how agents will respond afterwards. Importantly, neither properly describes the effect that I have tried to capture in this paper.

To see how, consider the difference between what we might call a state of "suspended belief," on the one hand, and a more firmly 'skeptical' (or perhaps, better, a 'cynical') outlook on the other. In a case

of suspended belief, I might take all contributions to a debate to be perfectly evidential (or 'informative'). It just so happens that on my current tally, the contributions for and against more or less balance each other out, leaving me without a clear reason to prefer either. So it's natural to describe me as being in a certain kind of 'disarray,' i.e., 'not knowing what to believe.' But it's important to note that I remain perfectly open to the possibility that further contributions could decisively tilt the balance in favor of one side or the other. So I remain 'tuned in,' precisely because I think there's real potential value to the signal that I might retrieve.

This is different from skepticism properly so-called: here, one doesn't just come out thinking that, as a matter of fact, my current balance of evidence fails to favor *p* or not-*p*. Instead, one comes out thinking that none of this is evidence at all, and that this whole domain of discourse is just bullshit: all noise and no signal.

In brief, whereas 'suspended belief' should motivate you to stay 'tuned in' to the channel, since you continue to think that it's transmitting a potentially valuable signal, Rini's 'generalized skepticism' should motivate you to 'tune out' altogether.

For an illustration, consider the public messaging regarding the health benefits of masking during the earlier stages of the covid-pandemic. 'Suspending belief' on this conflicting 'expert testimony' is fully compatible with continuing to think that these people are all genuine experts. Presumably, the reason they disagree is just that it's genuinely hard to know the truth about these things. Nonetheless, we should stay tuned in, because there's real value to their contributions, and hopefully, soon enough, we will see an evidence-based convergence of opinion in one direction or the other.

By contrast, the skeptic might say that regardless of their scientific credentials, none of these people are actually speaking from knowledge: instead, they are all engaged in virtue-signalling, hoping to gain the unlikely five minutes of fame that this crisis provides them.

But clearly, these are not the only two options. A third response, which I take to be at least as common, would go something like this: you quickly settle on what you take to be the correct view, and likewise settle into a groove where you primarily communicate with others who share that view. Soon enough, you will come to think that people holding the opposing view are not just epistemically incompetent but also perhaps morally corrupt. (I.e., they don't have the right values: they value their own minor personal freedoms over the health of vulnerable populations; or they are 'sheeple,' only too happy to let governmental organizations use a public health crisis as a pretext for grabbing more power.)

It is this third type of option that I think is missing from standard accounts of FN, but which I think is a common and predictable effect of it. Because, clearly, many of us don't change our minds in response

to FN. That is, we don't suddenly acquire new (false) beliefs as a result of exposure to FN. Nor do we abandon our pre-existing beliefs because we realize that our evidential situation is now suddenly so much more complicated.

Were we just lucky to be immune to FN? Or were we perhaps less gullible, more epistemically rational than others? Should we think that we suffer no bad effects as a result? This seems naïve. Even if FN doesn't have a direct impact on our doxastic outlooks, it still impacts the quality of our information environment, and thereby, over time, will have a detrimental effect on our political agency.

By altering the signal-to-noise ratio in our information economy at large, FN can change the incentives for engaging in political discourse on broad and inclusive platforms, thereby effectively forcing information sharing and joint deliberation 'underground.' That is, the prevalence of FN motivates us to seek out information sharing platforms or networks where we can expect a higher signal-to-noise ratio. These networks will naturally be shaped by pre-existing social bonds of epistemic trust and will accordingly gravitate toward those who already share our political outlook and sense of value. In other words, FN moves us in the direction of 'enclave deliberation,' and perhaps ultimately to 'echo chamber' formation, i.e., a socio-epistemic structure where certain kinds of perspectives and sources of information have been taken out of the reckoning altogether, perhaps by way of the move that I have elsewhere dubbed 'evidential preemption' (Begby 2021*b*).[21]

From one point of view, this might well seem like an epistemic net benefit: isn't it a good thing to filter out all the noise so that we may better focus on properly credentialed opinions that are more likely to be true? Maybe. But once we think of it in terms of the epistemology of democracy, and of collective decision-making more broadly, the problem takes a different hue, since we have to assume that exactly the same reasoning will also occur on the other side of the political spectrum. In fact, that's exactly how echo chambering predictably leads to 'belief polarization,' as a result of which we become – on all sides of the political spectrum – *more*, not less convinced that we are right and the others wrong. And precisely because they are wrong, we don't have to take their views into serious account, thereby artificially reinforcing the cycle.[22]

[21] To be clear, the point here is not (simply) that echo chamber formation reduces the likelihood that we will ever be *exposed* these other perspectives and sources of information, but rather that it's common ground within our peer group (our 'enclave') that such views are not to be taken seriously. Even so, it might be beneficial for us to retain *some* level of exposure to such views. For one, it might be strategically to keep tabs on what 'the others' are likely thinking. But also, and frankly, periodic reminders that this is indeed what the others are thinking might also be vital to maintaining cohesion in our own peer group over time.

[22] Drawing on competing accounts of the socio-political processes leading up to the 2016 US election, Nguyen (2021) contrasts a 'polarization' account and a 'propaganda' account, one key difference being that while the former is

The result is, I think, recognizably consistent with us, (i) retaining our political beliefs – in fact maybe even strengthening them – but also (ii) wielding a predictably corrosive effect on our shared political environment and on our willingness to engage in open-minded, maximally inclusive deliberation, informed by a shared sense of value. In other words, it undermines democracy's ability to realize its epistemic potential, thereby contributing to undermining our political agency.

Importantly, this outcome has nothing in particular to do with individual epistemic rationality, as standard analyses of FN effectively presuppose. This would be a natural hypothesis, of course, if we continued to suppose that the point of FN is to target our first-order beliefs or credences. But once we come around to thinking that its aim is rather to divide us as a political community, we get a very different view as a result: now, the *collective predicament* of belief polarization can be seen as a predictable result precisely of individuals trying to implement relevant epistemic norms in what they reasonably take to be an antagonistic information environment.[23]

Moreover, I would suggest that their taking it to be so in no small part a result of high-profile discourse, with accompanying public awareness campaigns, extolling the dangers of an 'FN crisis' or 'epidemic.' This discourse implicitly encourages us to think of others as epistemically incompetent and as potential 'disease vectors' in the spreading of disinformation, and therefore to be very selective about who we trust and who we sincerely engage with in forming our own political beliefs and attitudes.

For an empirical perspective on this, consider Altay and Acerbi's recent finding (2023: 1) that

> the strongest, and most reliable, predictor of perceived danger of misinformation is the third-person effect (i.e., the perception of that others are more vulnerable to misinformation than the self) and, in particular, the belief that "distant" others (as opposed to family and friends) are vulnerable to misinformation.

symmetrical – i.e., occurs spontaneously but predictably on both sides of the political spectrum – there is reason to believe that one side is in fact more susceptible to propaganda than the other. After reviewing the evidence, Nguyen cautiously favors the propaganda account. But this is not a challenge to the view offered here, as my argument cuts across the categories that Nguyen considers. Notice that Nguyen (2021: 173) defines 'propaganda' in terms of agents filling 'some part of the media environment with false or misleading information, for political purposes.' On this definition, it should be clear that FN also qualifies as propaganda, and part of my argument is simply that we should understand propaganda as an *accelerant* to processes of belief polarization that might already be in operation. But more centrally, Nguyen's survey follows broader trends in the literature in considering propaganda primarily in terms of its direct doxastic effects – i.e., its persuasive ability – and argues that one side of the political spectrum has shown itself to be more susceptible to propagandistic persuasion. By contrast, my argument brings to light to an important *indirect* effect of FN, namely its power to bring about changes in our patterns of deliberative interaction – thereby contributing to belief polarization – even without persuading us of its content. Nothing in Nguyen's argument, or in the broader empirical literature that I have surveyed, militates against thinking that this is a symmetrical development, one we should expect to occur on both sides of the political spectrum and in roughly equal measure.

[23] Cf. Begby 2022.

Differently put: individual concerns about FN is less the concern that *I* (or people in my peer group) might get duped as it is the concern that 'distant' (or 'outgroup') others will. By whipping up a 'moral panic' around the effects of FN, the primary way that these awareness campaigns may succeed in modifying *my* behavior is by encouraging me to increase my epistemic distance to these others, thereby creating conditions ripe for belief polarization and echo-chambering. My confidence in my own (and my peer group's) ability to process information stays put, while my confidence in *their* abilities goes down. Since this effect is presumptively symmetrical (i.e., occurs on all sides of the political spectrum), it has predictably deleterious consequences for the epistemology of democratic decision-making, which fully relies, at least according to Democratic Optimists, on our willingness and ability to engage in information sharing and joint deliberation on the most inclusive platforms.

6. Why does Fake News thrive on social media?

This angle might also help answer one of the most frequently asked questions in the literature, namely *why does FN thrive on social media, in particular*? (Clearly there was FN before the internet-era. Nonetheless, there seems to be wide agreement that there's something about social media that seems particularly vulnerable to being overrun by FN.)

Again, typical answers focus on the persuasive power of FN, leading us to speculate about why people seem so much more gullible in the context of social media. Is it simply that social media brings out the worst in us, i.e., just 'makes us stupid'? Or is it perhaps because, as Rini (2017) argues, the peer-to-peer structure of social media interactions succeeds in mimicking ordinary, everyday testimony to a sufficient degree that people drop their critical guard, becoming more trusting than they would be of political discourse occurring on any other platform?

Again, I think that focusing on individual cognitive behaviors may be a distraction; instead, FN thrives on social media for exactly the same reason that we might once have hoped that ground-level democratic processes could thrive on social media.

To see how, consider a typically optimistic perspective on internet and social media from some 20-25 years ago. Question: might social media be an important factor in encouraging political engagement and strengthening our ground-level democratic processes? Answer: Yes! Everyone knows that democracy has a participation problem. Democracy works best when it works in a maximally inclusive way, where, so far as is possible, everyone's opinion gets thrown into the mix. As such, it depends on high-fidelity, low-barrier channels of communication and platforms for joint deliberation. For democracy to really deliver on its

epistemic promise, we need access to broadly inclusive, decentralized peer-to-peer networks, where political opinions can be offered and debated in real-time, with very low barriers to participation. And that's precisely what social media promise to provide.

In hindsight, this may seem like naive techno-optimism. But it's worth asking, what has happened in the meantime. For notice how the same mechanism that would make social media potentially promising for the epistemology of democracy would also makes it (i) an important target for subversive campaigns such as FN, (ii) particularly vulnerable to such campaigns, precisely because the barrier is so low.

We can add that the essentially decentralized character of social media makes monitoring for information quality very difficult. We could, of course, try to centralize quality-control, as some social media platforms have eventually sought to do, for instance in the form of fact-checking. But imposing such centralized fact-checking will also, at the same time, detract from its value as a platform for maximally inclusive political deliberation.[24] (To say nothing of the actual *quality* of such fact-checking,[25] or whether the fact-checkers are likely to be *widely seen* as objective or unbiased in a political environment that is already significantly polarized.[26] It's a familiar dynamic that the perceived political bias of such institutionalized constraints will likely just push partisans 'underground' to other platforms, once again reinforcing the process of belief polarization.)

With this in mind, consider different strategies that subversive agents might think to deploy if they were looking to disrupt baseline political processes in the social-media era. They could engage in a select campaign of high-quality disinformation, hoping to acquire a broad readership via sharing and liking, and eventually to trick some people into adopting the relevant false beliefs. Or they could do a perfunctory cost-benefit analysis and instead decide to pump out vast amounts of low-quality disinformation so as to reduce the signal-to-noise ratio, thereby hoping to fan processes of echo-chambering and belief polarization.

As I've argued, this low-quality FN might not be fit to persuade anyone. But that's also not it's job. Instead, it has the benefit of being cheap and easy, and can plausibly succeed in a nearby aim: simply to disrupt the easy opportunities for high-quality political engagement that social media seemed to provide.

[24] Cf. Stewart 2021.
[25] Cf. Winsberg 2023; Fernandez-Roldan and Teira 2024.
[26] Cf. Gibbons 2023.

7. Conclusions

This paper has argued that we should seek to understand the FN-phenomenon – and to understand the relative prevalence of low-quality FN in particular – as a simple and cost-effective subversive intervention on ground-level democratic processes of information sharing and public deliberation. In particular, I have argued that FN need not be designed for *persuasion* in order to achieve this aim: what matters is just that there's enough of it to significantly reduce the signal-to-noise ratio on our most inclusive communication channels. For these purposes, quantity trumps quality. Hence, the relative prevalence of low-quality FN: there is simply no need for a better product.

Adopting this perspective gives us a new handle on FN as an empirical phenomenon, but also draws into doubt the value, relevance, and efficacy of various kinds of 'public awareness' campaigns designed to protect us against FN. These kinds of campaigns very much suppose that the point of FN is to instill persuasion (or minimally doubt). But as we have focused on building individual resistance against the persuasive power of FN, we may well have lost sight of this other politically corrosive effect that FN might have: that it motivates us to disengage from the kind of maximally inclusive information sharing and joint deliberation processes that democratic theorists would have us aspire to, leading ultimately toward echo chamber formation and resultant belief polarization.

While these are regularly taken to undermine the epistemic potential of democracy, we don't have to suppose widespread epistemic irrationality at the individual level to understand how they might arise. Instead, they might well be seen as arising from a reasonable concern with personal 'epistemic hygiene' (reasonable, not least, in light of these public awareness campaigns). In the end, it is no small irony that the whole discourse centered on creating public awareness of a 'Fake News crisis' or 'Fake News epidemic' might well have played some part in bringing it about. Such discourse entrenches the idea that a significant portion of our fellow citizens are potential 'disease vectors' in the spreading of disinformation, and hence that we would do better, epistemically speaking, by significantly shrinking our 'epistemic bubble' – i.e., the people in whom we regularly place our epistemic trust for the purposes of information sharing and joint deliberation – to the point of its becoming a structurally reinforced 'echo chamber.'[27]

[27] Work on this article was supported by SSHRC Insight Grant 435-0222-0193. For discussion and important input along the way, I owe thanks to Glenn Anderau, Holly Andersen, Jennifer Saul, audiences at the APA Public Session on the Philosophy of Fake News (Vancouver, April 2019), Technical University of Berlin (May 2023), University of Oslo (May 2023), University of Indiana, Bloomington (September 2023), the Western Canadian Philosophical Association (October 2023), as well as students in my Fall 2023 Political Epistemology course at Simon Fraser University.

References:

Anderau, Glenn 2021. 'Defining Fake News' *Kriterion* 35(3): 197-215.

Anderau, Glenn 2023. 'Fake News and Epistemic Flooding,' *Synthese* 202(106).

Anderson, Elizabeth 2006. 'The Epistemology of Democracy,' *Episteme* 3(1-2): 8-22.

Altay, Sacha, and Alberto Acerbi 2023. 'People Believe Misinformation is a Threat because they assume Others are Gullible,' *New Media & Society*, Online First.

Altay, Sacha, Manon Berriche, and Alberto Acerbi 2023. 'Misinformation on Misinformation: Conceptual and Methodological Challenges,' *Social Media + Society* January-March 2023: 1-13.

Begby, Endre 2021a. *Prejudice: A Study in Non-Ideal Epistemology*. Oxford University Press.

Begby, Endre 2021b. 'Evidential Preemption,' *Philosophy and Phenomenological Research* 102(3): 515-530.

Begby, Endre 2022. 'From Belief Polarization to Echo Chambers: A Rationalizing Account,' *Episteme* 21(2): 519-539.

Blake-Turner, Christopher 2020. 'Fake News, Relevant Alternatives, and the Degradation of our Epistemic Environment,' *Inquiry*, DOI: 10.1080/0020174X.2020.1725623

Brennan, Jason 2016. *Against Democracy*. Princeton, NJ: Princeton University Press.

Buzzell, Andrew and Regina Rini, 2022. 'Doing your Own Research and Other Impossible Acts of Epistemic Superheroism,' *Philosophical Psychology* 36(5): 906-930.

Dentith, M. R. X. 2017. 'The Problem of Fake News,' *Synthese* 8(1-2): 65-79.

Dutilh Novaes, Catarina, and Jeroen de Ridder 2021. 'Is Fake News Old News?' In S. Bernecker, A.K. Flowerree and T. Grundmann (eds.), *The Epistemology of Fake News*. Oxford University Press.

Ecker, Ullrich, Jon Roozenbeek, Sander van der Linden, Li Qian Tay, John Cook, Naomi Oreskes, and Stephan Lewandowsky 2024. 'Misinformation Remains a Threat to Democracy,' *Nature* 630: 29-32.

Fernandez-Roldan, Alejandro, and David Teira 2024. 'The Epistemic Status of Reproducibility in Political Fact-Checking,' *European Journal of Philosophy of Science* 14(12). https://doi.org/10.1007/s13194-024-00575-8

Gelfert, Axel 2018. 'Fake News: A Definition,' *Informal Logic* 38(1): 84-117.

Gibbons, Adam F 2023. 'Bullshit in Politics Pays,' *Episteme* FirstView.

Guess, Andrew M., Brendan Nyhan, and Jason Reifler 2023. 'Exposure to Untrustworthy Websites in the 2016 U.S. Election,' *Nature Human Behavior* 7:1096–1105.

Herley, Cormac 2012. 'Why Do Nigerian Scammers say They Are from Nigeria?' *Microsoft Working Papers*. https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/WhyFromNigeria.pdf

Jaster, Romy, and David Lanius 2021. 'Speaking of Fake News: Definitions and Dimensions,' in S. Bernecker, A.K. Floweree, and T. Grundmann (eds.), *The Epistemology of Fake News*. Oxford University Press.

Landemore, Helene 2012. *Democratic Reason*. Princeton, NJ: Princeton University Press.

Levine, Timothy R. 2022. 'Truth-Default Theory and the Psychology of Lying and Deception Detection,' *Current Opinion in Psychology* 47: 101380.

Lynch, Michael 2016. 'Fake News and the Internet Shell Game,' *New York Times*. https://www.nytimes.com/2016/11/28/opinion/fake-news-and-the-internet-shell-game.html

Nelson, Jacob L., and Harsh Taneja 2018. 'The Small, Disloyal Fake News Audience: The Role of Audience Availability in Fake News Consumption,' *New Media & Society* 20(10): 3720-3737.

Nguyen, C. Thi 2020. 'Echo Chambers and Epistemic Bubbles,' *Episteme* 17(2): 141-161.

Nguyen, C. Thi 2021. 'Was it Polarization or Propaganda?' *Journal of Philosophical Research* 46: 173-191.

"'Nigerian prince' email scams still rake in over $700,000 a year—here's how to protect yourself." *CNBC*. https://www.cnbc.com/2019/04/18/nigerian-prince-scams-still-rake-in-over-700000-dollars-a-year.html

Rini, Regina 2017. 'Fake News and Partisan Epistemology,' *Kennedy Institute of Ethics Journal* 27(S2): 43-64.

Rini, Regina 2021. 'Weaponized Skepticism: An Analysis of Social Media Deception as Applied Political Epistemology,' in E. Edenberg and M. Hannon, (eds)., *Political Epistemology*. Oxford University Press.

Schramm, Wilbur 1954. 'How Communication Works,' in W. Schramm, ed., *The Process and Effects of Mass Communication*. Urbana, Il: The University of Illinois Press.

Stewart, Elizabeth 2023. 'Detecting Fake News: Two Problems for Content Moderation,' *Philosophy & Technology* 34:923–940.

Sunstein, Cass 2002. 'The Law of Group Polarization,' *Journal of Political Philosophy* 10(2): 175-195.

Talisse, Robert 2021. 'Problems of Polarization,' in E. Edenberg and M. Hannon, eds., *Political Epistemology*. Oxford University Press.

'Ten Questions for Fake News Detection' 2016. *The News Literacy Project*. https://guides.lib.jjay.cuny.edu/ld.php?content_id=28546482

Wagner, Marie Celeste, and Pablo J. Boczkowski 2019. 'The Reception of Fake News: the Interpretations and Practices that Shape the Consumption of Perceived Misinformation,' *Digital Journalism* 7(7): 870-885.

Winsberg, Eric 2023. 'We Need Scientific Dissidents now more than Ever,' *The Chronicle of Higher Education*, August 10, 2023.

World Economic Forum 2024. *The Global Risks Report*. 19th edition. https://www3.weforum.org/docs/WEF_The_Global_Risks_Report_2024.pdf