THE FILTER MAP:

Media and the Pursuit of Truth and Legitimacy

DEEN FREELON
Associate Professor
School of Media and Journalism
University of North Carolina
What kinds of media messages should we choose to receive, and how seriously should we take them? These questions are much more urgent now than they were in the pre-digital era, when we had far less control over the information we consumed. Back then, nearly all our media options came from the major TV networks, movie studios, book publishers, radio broadcasters and local news monopolies. Today, anyone can add to our inexhaustible menu of media choices, which continue to expand without pause. With so many choices at our fingertips, filtering becomes a top priority to arrive at any coherent understanding of the world beyond our direct experience. But what principles ought to drive our filtering and interpretive criteria?

The standard answer to this question starts from the assumption that many of us don’t always make the best choices about what information to consume. As with food, what appeals to us most immediately is not always the healthiest or most useful — for example, soft news and infotainment are sometimes blamed for emphasizing the most salacious and least consequential aspects of public affairs. But even worse is factually untrue content designed to look authentic (what is sometimes called “fake news”), which usually targets those who agree with the false message. Extreme opinions can be equally harmful, as when they advocate for the systematic injury, oppression, or extermination of entire ethnicities, sexual orientations, or religious groups. I argue that, generally, people are well justified in avoiding such content.

It’s easy to come up with examples of messages we do or don’t want to encounter. What has proven more difficult is specifying generalizable criteria for what makes a given piece of content worthy of our attention and consideration. One popular solution holds that we need to configure our filters to include messages “from the other side;” in other words, that lie outside our narrow domain of direct interest and opinion. In his recent book #Republic, Cass Sunstein advocates for media systems that give their audiences “unplanned” and “unwanted” experiences, apparently assuming that all exposure to other citizens’ opinions is inherently good. Clay Johnson, author of The Information Diet, cautions against consuming too much “mass affirmation” (i.e. content from sources we agree with), likening its mental effects to those of refined sugar on the body. And Eli Pariser, who gave us the now-household term “filter bubble,” points out that when services like Google and Amazon automatically show us what they think we want to see, they prevent us from seeing lots of potentially useful stuff.

These insights give us a helpful start on our quest for better information filters, but they don’t go far enough. As a social scientist, I need to raise a slight empirical quibble with the notion that most people only consume content they agree with. Careful research has shown this to be incorrect, instead finding that while people prefer messages that support their

---

1 If your household includes media professionals, that is.
opinions, they do not systematically avoid disagreeable messages. So while media thinkers might wish for most media diets to be perfectly ideologically balanced at 50/50, the true balance is almost never 100 percent one-sided. But aside from that, I object more fundamentally to the notion that all mass affirmation (or “opinion congruency,” as scholars sometimes call it) is always bad, and its corollary that unwanted or unplanned encounters are always good. In true academic fashion, I will argue that it depends: each can sometimes be good and sometimes bad. This implies that we should not program our filters to use agreeableness as a primary decision rule, nor should we use it to determine whether to take messages seriously.

I have an alternative suggestion for how to think about the kinds of information that are more and less worthy of our time. It involves not one conceptual dimension but three, which combine to create a filter map onto which we can project most of the messages we will encounter. The first is agreeableness, which captures the Sunsteinian/Johnsonian/Pariserian concept of aligning with our preexisting opinions. The second, truth value, is easy to define in principle but hard to apply in practice: it is simply whether a given message is true or false. To get a sense of how fraught this distinction can get, simply page back through some of public debates around the professional fact-checking industry from the past few years. The third axis I call legitimacy, and it is perhaps the most difficult to define of all. The term is borrowed from Daniel Hallin’s 1986 book The Uncensored War, in which he coins the phrase “sphere of legitimate controversy” to cover all opinions considered to be acceptable within a given society. In the United States, as in advanced democracies more generally, this usually reduces to an opinion’s adherence to widely accepted ethical norms like freedom, equality, fairness and human rights. While there is considerable debate around what kinds of opinions comport with such principles, it can be safely said that crimes like racial discrimination, torture and arbitrary detention definitively violate them. Legitimacy is solely a property of opinions, while truth value applies only to facts. Agreeableness cuts across both facts and opinions, as I’ll show a bit later.

The filter map that emerges from the intersection of these three dimensions (Table 1) offers a simple guide to the kinds of information one should strive to consume and avoid. Ideally, our media filters would optimize for truth and legitimacy, ensuring that both agreeable and disagreeable content and sources are included (the map’s four blue cells). By the same token, false and illegitimate messages would be excluded, again regardless of agreeableness.


(the four white cells). The conceptual leap I make here is from considering disagreeableness as a virtue in itself, as the previously mentioned authors have done, to distinguishing between more and less desirable types of disagreeable content. There are many claims and opinions we should rightly dismiss out of hand, but there are others we should entertain despite disagreeing with them. This essay explains why I think filtering information this way is beneficial, illustrating key points primarily by example. I first discuss agreeableness in detail, next proceeding to truth value and then legitimacy, covering all eight of the filter map’s cells along the way.

**TABLE 1: THE FILTER MAP**

<table>
<thead>
<tr>
<th></th>
<th>AGREEABLE</th>
<th>DISAGREEABLE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CLAIMS OF FACT</strong></td>
<td>True</td>
<td>Facts that support your case</td>
</tr>
<tr>
<td></td>
<td>False</td>
<td>Comforting lies</td>
</tr>
<tr>
<td><strong>OPINIONS</strong></td>
<td>Legitimate</td>
<td>Opinions you agree with and are compatible with broadly shared values</td>
</tr>
<tr>
<td></td>
<td>Illegitimate</td>
<td>Opinions you agree with and are incompatible with broadly shared values (e.g. your prejudices)</td>
</tr>
</tbody>
</table>

**Agreeableness**

We all have opinions, and most people don’t particularly enjoy having them threatened or belittled. In the pre-digital era, most people formed most of their political opinions based on content produced by a small collection of tightly-controlled media outlets. Today, our opinions can be informed by nearly any perspective imaginable. Given this situation, it is little wonder that many tend toward news and information that consistently flatters their preexisting opinions. Such “hyperpartisan” content mostly reaches the politically engaged, while apolitical citizens usually try to filter politics out of their personalized information streams as much as possible. Either way, people generally seek content that fits their interests, which often means opting into information streams that are low on disagreeableness.

Of course, this isn’t inherently bad; on the contrary, it’s a straightforward consequence of using filters to perceive our way through a media-saturated world. But our filters aren’t perfect, and most of us end up encountering disagreeable content on a fairly regular basis. When something gets through that we disagree with, we have two basic choices for how to proceed. We can shoo it away as quickly as possible, as we might an errant housefly; or we can sit with it awhile, consider its strengths and weaknesses, and perhaps even allow it to shift our thinking, if only a little. But the bare fact that a message conflicts with our opinions is not enough information to decide which of those paths to take.
A better way of doing so involves knowing whether a message is a claim of fact or an opinion. To refer to something as a “claim of fact” is to assert that it is falsifiable, meaning capable of being proven true or false. The statements “the earth is round” and “the earth is flat” are both falsifiable, but of course one is true while the other is false. In contrast, the statement “the earth is beautiful” is not falsifiable because it expresses an opinion and is therefore not subject to empirical verification. In the simplest analysis, fact claims are either true or false, although fact-checkers like PolitiFact have introduced graded spectra of truthfulness to capture the slippery ambiguity of contemporary political speech. The analogous categories for opinions are legitimate and illegitimate, which I’ll explain at greater length later.

Truth value

I want to cover truth value first because it’s more familiar and therefore easier to grasp than legitimacy. Epistemologists have debated the question of what exactly constitutes truth since the time of Socrates, but that’s not our present concern. All that’s required to proceed is a fundamental faith in empiricism, the ability to determine what is and isn’t true by observing the world. If we can assume that’s theoretically possible for all fact claims, and practical for many important ones, we can start to consider what kinds of claims we want to encounter and taking seriously. The answer here is deceptively simple: we should embrace truth and reject falsehood. Yet in practice this turns out to be much more difficult than it appears.

One of the most powerful causes of this difficulty is the agreeableness dimension. This is because it’s much easier to decide whether a message is agreeable than whether it is true. Indeed, it’s so much easier that some people, including the current U.S. president, have succumbed to the temptation to erase all distinctions between the two. According to such individuals, what is agreeable is true, and what is disagreeable is false — this is how “fake news” comes to mean “messages that contradict my preferred fact claims.” Only agreeable fact claims matter, this perspective dictates, and everything else can be dismissed with the freshly-redefined label of “fake.” The Oxford English Dictionary saw fit to recognize this epistemological chimera with its 2016 Word of the Year, “post-truth.”

I’ll give the post-truthers this much: agreeable facts are fine, so long as they are, in fact, facts. As much as I wish this went without saying, I’ll say it again: what should matter is true vs. false, not agreeable vs. disagreeable, because falsehoods outside of fictional contexts are generally bad. Using agreeableness as a primary information filter causes two kinds of problems specific to fact claims: first, it increases users’ vulnerability to seductive falsehoods; and second, it lowers the likelihood that users will encounter true but disagreeable messages. Lies designed to look true are the other kind of fake news, and American pranksters, Macedonian teenagers, and Russian propagandists (among others) have been known to spread them. One of the most effective techniques to boost uptake of false messages is to appeal to a political faction’s core beliefs. For example, “Pope Francis Endorses Trump”
gained currency among the right in 2016, while “Florida School Shooter Tied to White Nationalist Group” appealed more to the left in early 2018. Given that confirmation bias predicts we will accept agreeable claims less critically, we should be vigilant about rejecting them emphatically whenever they are debunked.

But congenial falsehoods are not the only hazard awaiting those with a distaste for disagreeableness. They will also miss inconvenient truths, and while that may keep one’s ego safe, it’s a dangerous way to build a worldview. To put it lightly, ignoring potentially important facts may result in extremely flawed decisions. Adherents of the flat earth conspiracy theory have failed to incorporate the abundant evidence of our planet’s spherical shape into their cosmology (although they accuse everyone else of doing the same with their preferred “evidence”). Somewhat more seriously, if Edgar Welch had paid more attention to the empirical doubts raised about a possible child sex ring housed underneath the Comet Ping Pong restaurant in Washington, D.C., he might not have fired a gun on the premises while trying to liberate its nonexistent victims. These examples and countless others stem from an unwillingness to perceive or engage with valid yet opinion-threatening facts. It is a tendency we should endeavor to eliminate.

The final category to consider in our truth-value quartet is disagreeable falsehoods. These are lies whose danger to ourselves and our belief systems allows us to identify them quickly and conclusively. As such, few of us have difficulty addressing them appropriately. Most fit the conceptual, if not the legal, definition of libel or slander, such as allegations that President Obama was born in Kenya (birtherism), or that members of the Bush administration had advance knowledge of or helped plan the 9/11 attacks (9/11 trutherism). When lies like these threaten our opinions, we typically waste little time dismissing or debunking them, as we should.

Legitimacy

Having completed our analysis of desirable and undesirable fact claims, let us now turn to the domain of opinion. Because opinions can’t be definitively judged based on empirical evidence, we must use alternative criteria to decide which to allow and deny our attention. Truth value is the only valid criterion for assessing fact claims, and the only widely-accepted equivalent for opinions is culturally-specific consensus. I capture this notion with the term “legitimacy.” In the West, legitimate opinions generally align with moral principles that are enshrined across a broad array of societies, governments, philosophies, and religions. Examples include freedom, justice, equality, fairness, universal human rights, due process, and non-maleficence (“do no harm”). While it may seem that some of these terms have been overused to the point of meaninglessness, I argue that they form a meaningful bulwark
against many of humanity’s worst impulses. For example, it would be difficult to argue that practices such as torture, arbitrary detention, theft, collective punishment, ethnic persecution, and sexual assault are “legitimate” under most widely-subscribed moral codes. Yet there are active schools of thought today that openly advocate for each of these.5

Under most circumstances, such opinions do not warrant serious consideration, or even attention. This is not because they are disagreeable, but rather because they are illegitimate. The distinction between the two is subtle but useful, and thus worth explaining. Agreeableness is a fundamentally subjective characteristic; what is agreeable to me may or may not be to you. In contrast, legitimacy is intersubjective, deriving its status as such from moral values shared across governments, cultures, religions, and philosophies. It is not universal (if the internet has taught us anything, it’s that nothing is), but its components are officially recognized by enough individuals and institutions to recommend its use as an opinion filtering criterion.

This means that opinions can simultaneously be agreeable and illegitimate. Bigotries fit this category, whether held against a specific ethnicity, gender identity, religion, economic class, disability status, etc. When held by the individual doing the content filtering, such prejudices are agreeable in the same sense as any other political opinion from their perspective. But they are intersubjectively illegitimate, which is why most bigots deny their bigotry even when confronted with undeniable evidence thereof. Recently we have begun to see shameless advocacy for prejudicial views in the mainstream media, with two recent examples involving questioning the personhood of Jews and denying the franchise to women.6 As illegitimate opinions, our framework would recommend they be dismissed out of hand for flouting the notions of human rights and equality that most Western citizens take for granted. Individuals holding such opinions should seriously reconsider them in light of the great harm they continue to cause.

Moving diagonally across the map, let’s next consider disagreeable yet legitimate views. These are opinions one disagrees with but that still lie within Hallin’s sphere of legitimate controversy. To reiterate, this is the set of political opinions within a society that may conflict but are considered reasonable to hold, and that the press and other major institutions thus take seriously. To categorize an opinion as disagreeable and legitimate is to acknowledge its political validity in spite of your personal rejection of it. You can see its merits (or at least understand people’s reasons for holding

5 This excludes arguments that such practices are something other than what they are widely considered to be, e.g. denials that waterboarding is torture. See Greenwood, Max. “Trump: Waterboarding Isn’t Torture.” Text. TheHill, January 26, 2017. http://thehill.com/policy/national-security/316435-trump-waterboarding-isnt-torture.

it), debate it in good faith, and perhaps even consider subscribing to it. In contrast, illegitimate opinions occupy what Hallin calls the “sphere of deviance,” which is reserved for ideas so diametrically opposed to the shared values I outline above that to even entertain them would cause intolerable harm. This is why we shouldn’t have serious national conversations about whether Jews are people or whether women ought to be able to vote.

The remaining two categories, agreeable/legitimate and disagreeable/illegitimate, are fairly straightforward. In most cases, our instincts will direct us to accept the former and reject the latter, as we should. Legitimate opinions are much easier to embrace when we also agree with them, and we are just as quick to condemn illegitimate opinions we disagree with.

I hope the above discussion amply demonstrates that, similarly to truth value, an optimal opinion filtering criterion would include both types of legitimate opinions while excluding both types of illegitimate ones. Legitimate dissent is essential to democratic practice in pluralistic societies — we can’t simply wall ourselves off from every opinion we disagree with, if only because some of those opinions will inevitably shape public policy. So we need to decide which of all the opinions we don’t hold that we will respect and which we will not, because not all opinions deserve respect. I believe the concept of legitimacy, imperfect as it is, offers a useful guide for making such distinctions.

Implementation and limitations

In this concluding section I will address the question of how the filter map might be implemented, as well as some potential objections to its key components. Developing this framework has been an engaging intellectual exercise, but if it can’t be applied effectively in practice, it will have been strictly an academic one. But several existing initiatives demonstrate functionality that points the way toward what I’ve proposed. Most social media platforms are built to maximize the agreeableness of the content the user sees, so most of those features would need to be discarded at the outset. We are looking to build a system that delivers true and legitimate content from across our personal agreement spectrum.

In some ways, this desired result resembles the content of a national daily newspaper or newscast — what Sunstein called a “general interest intermediary.” But our ideal system would improve upon the well-known legacy news outlets in several ways. First, it would offer a much wider range of content than what has traditionally been considered “newsworthy” by drawing on the value judgments of more than just news editors about what is worth sharing. This is extremely important given the longstanding problems with institutional inclusion in journalism — whenever news judgments come from a largely homogeneous group, certain topics will inevitably be underrepresented. Relatedly, some of the most deeply-ingrained journalistic practices require the inclusion of content that lies outside the filter map’s blue cells. When reporters insist on presenting views from “both sides” of certain issues, falsehoods are sometimes presented as equivalent to truth.
This used to be common in discussions of climate change, where climate deniers were presented on equal footing with the scientific consensus, although that has changed for the better recently. The desire to appear objective has also left news outlets poorly equipped to handle situations where one side lies disproportionately more often than the other. A filtering system that does not claim objectivity would not have this problem: all messages that meet our blue criteria would be admitted. Instead, it would seek a different kind of balance — that between agreeable and disagreeable blue content.

Most people agree that truth is better than falsehood, but distinguishing between the two seems to get more difficult by the day. If my social media feeds allowed me to reliably filter out all false content, I’d flip those switches right now. In the absence of such a magic bullet, how else might we cope with all the fakery out there? No single approach will solve the problem by itself, but multiple technical and individual interventions can move us forward from where we are. I am encouraged by Facebook’s practice of attaching “related articles” to content that has been flagged by third-party fact-checking organizations as false. This feature results in far lower rates of false content sharing than an earlier one that added large “Disputed” badges to such content.7 Google has increased efforts to remove misleading content from its search results by demoting the sites that carry it in search results8 and preventing them from masking their countries of origin.9 Browser extensions like Newscracker and FactoidL10 attempt to project a similar logic onto the entire web, and although they currently suffer from substantial flaws, future versions might work well enough for daily general use.

Stamping out flagrant falsehoods is hard enough, but what about messages from that gray area in the middle of the truth value spectrum? Most content doesn’t come in the form of neatly true or false statements like the ones from epistemology textbooks — many are embedded in complex messages that can’t easily be labeled one way or the other. An illustrative case in point is PolitiFact’s 2011 Lie of the Year, which was widely criticized by commentators on the left. That year, the organization bestowed the dubious distinction on Democratic politicians’ claims that Republicans had voted to “end” or “terminate” Medicare. PolitiFact described what the GOP had actually done thus: “The plan kept Medicare intact for people 55 or older, but dramatically changed the program for everyone else by privatizing it and providing

government subsidies.” Liberals sharply criticized PolitiFact’s decision to label their description of this move as a “lie,” let alone the most egregious of the year. The validity of that judgment turns on whether you believe changing Medicare into a different program with the same name (which PolitiFact acknowledged the GOP vote tried to do) constitutes “ending” it. This example illustrates that calling out falsehoods isn’t always easy, even for the experts. In such cases, a good filter map would include both perspectives so that users could make up their own minds.

As difficult as the domain of truth value can be sometimes, legitimacy is even more fraught. The shared values that are supposed to bound legitimate opinions are slippery and vague. Fortunately, there are many areas of moral consensus in our society, e.g. that slavery, murder, torture, rape, political violence, and other serious crimes should not be openly defended. Indeed, we rarely see straightforward advocacy of such crimes: we more often encounter attempts to re-categorize them as virtues or necessities. This is how waterboarding comes to be defined not as torture but as “enhanced interrogation,” and open racism rebrands itself from the shaved heads and combat boots of neo-Nazism to the suits and ties of the “alt-right.” News outlets’ editorial practices offer a model for rejecting such subtle infiltrations. There’s a reason NBC never hosts Islamic State members on Meet the Press, and it’s the same reason the Associated Press’ guidelines direct its reporters to identify the alt-right as the white supremacists they are. News outlets should obviously report on illegitimate beliefs when circumstances warrant; for example, when influential people proclaim them. But their stories should take pains to unambiguously separate such beliefs from those that deserve evenhanded coverage. Clearly-stated policies on how to handle known illegitimate philosophies can assist citizens as they go about deciding how seriously to take them.

Unfortunately, not all opinions are as clear cut. Nearly everyone claims to be in favor of freedom, justice, equality, human rights, etc., but a host of contradictory opinions come labeled with those terms. Some claim that abortion is a human right, while others consider it morally equivalent to slavery. Affirmative action has been called both an essential corrective to historical discrimination and an unacceptable infringement of racial and gender equality. Contested policies like these have no definitive solutions and can only be resolved within predefined ideological frameworks. Users might look to see what trusted and respected organizations such as the Southern Poverty Law Center, the Anti-Defamation League, and the NAACP have said about controversial opinions and movements. Blocking specific


individuals and outlets might prevent some illegitimate opinions from being displayed, but others will inevitably seep through. Users will need to interpret these individually based on their own individual senses of what is and is not legitimate.

The purpose of this essay has been to propose a rough but useful heuristic for deciding what kinds of messages we should want to encounter in our everyday mediated lives. It is not possible in this short space to address all possible objections, or to fully explain how it all might be implemented on a technical level. I leave those tasks to anyone who has been inspired to build on the ideas presented here. In closing, it’s important to remember that everything we accept as true or legitimate today is subject to change down the road. Therefore, it will sometimes be necessary to repudiate previously accepted messages that turn out to be false or illegitimate. It is this habit, more than any technical solution, that will bring us closer to the truth and rectitude most of us claim to want.

AUTHOR BIOGRAPHY

Deen Freelon is an associate professor in the School of Media and Journalism at the University of North Carolina at Chapel Hill. His research covers two major areas of scholarship: 1) political expression through digital media and 2) data science and computational methods for analyzing large digital datasets. He has authored or co-authored more than 30 journal articles, book chapters and public reports, in addition to co-editing one scholarly book. He has served as principal investigator on grants from the Knight Foundation, the Spencer Foundation and the U.S. Institute of Peace. He has written research-grade software to calculate intercoder reliability for content analysis (ReCal), analyze large-scale network data from social media (TSM), and collect data from Facebook (fb_scrape_public). He formerly taught at American University in Washington, D.C.