
Les fausses nouvelles : le nouveau visage d'un vieux problème

**Documents pour le séminaire du
jeudi, 5 octobre 2017 au
Carrefour des sciences, Université de
Montréal, Pavillon Lionel-Groulx,
3150, Jean-Brillant**

On the Study of Popular Subjects or Confessions of a Fake News Scholar

Tommaso VENTURINI

In these pages, for reasons that will try to make clear, I will not talk about fake news. I will, instead, talk about the *study of fake news*. And I will do my best to distinguish as sharply as possible between the two for I believe that, if salvation is possible, it can only be found in this distinction. ‘Salvation’ refers here to the reanimation of public debate in modern democracies, but also (and very distinctively) to our redemption as scholars studying fake news.

Truth be told, I have a love-hate relationship with fake news. While I feel uncomfortable when people treat me as expert on the subject, I have never been so popular than I after I started working on it. When, in March 2017, Liliana Bounegru, Jonathan Gray and I launched the *Fake News Field Guide* as the first project of the nascent *Public Data Lab* (I’ll come back to this initiative later), we could not imagine that we would raise so much attention. In just a few months, we received dozen offers of help by researchers and professionals from all over the world (so many in fact that we had problems in answering them). Several organisations invited us to use their database and technologies. Dozens of journalists, from all sorts of media outlets (included the most prestigious one), asked our opinions on the subject.

There is a whole tradition of artists feeling overly celebrated for works that they found too mainstream and neglected for what they consider their real contributions. Now I know that this feeling exists for scholars as well.

Five reasons NOT to talk about fake news

I do not want to sound snobbish, I do enjoy the reputation of expert. I just believe that this reputation is undeserved when it comes to fake news. I am not trying to be modest either. It is not just me – no one, I believe, can honestly claim to be a ‘fake news expert’, for the simple reason that there is no serious way to demarcate such a field of expertise. On the contrary, consensus is emerging among the scholars working on the subject that we should stop using this notion and start actively denouncing it. In the literature, I have found at least five reasons why, strictly speaking, there is no such thing as ‘fake news’ – five reasons why this paper is not about fake news.

(1) The first reason has to do with the awful vagueness of the term. A recent study by Tandoc, Lim & Ling (2017) has reviewed the definitions of ‘fake news’ from 34 academic papers published between 2003 and 2017 and found out that the authors used the term to refer to phenomena as different as satire, parody, fabrication, manipulation, native advertising, and propaganda. Ethan Zuckerman, director of MIT Center for Civic Media, in a post entitled *Stop saying “fake news”. It’s not helping* (2017), accuses the notion of being

“a vague and ambiguous term that spans everything from false balance (actual news that doesn’t deserve our attention), propaganda (weaponized speech designed to support one party over another) and disinformazya (information designed to sow doubt and increase mistrust in institutions)” (Zuckerman, 2017).

(2) Because of its vagueness, the term ‘fake news’ is not only scientifically shaky, but also politically dangerous, because it lends itself to be used as a weapon to discredit opposing sources of information (US President Donald Trump has provided several excellent examples of such use). Since there is not precise way of demarcating fake news, the term becomes nothing more than a rhetorical accusation. Claire Ward, research director of *First Draft* (a non-profit coalition bringing together the most important journalism and media platforms - firstdraftnews.com) writes in a report on disinformation for the Council of Europe:

“Another reason why the term “fake news” is insufficient and dangerous to use is because it has been appropriated by politicians around the world to describe news organisations whose coverage they find to be problematic. The term “fake news” is being used as a mechanism for clamping down on the free press, and serves to undermine trust in media institutions, hoping to create a situation whereby those in power can circumvent the press and reach supporters directly through social media. It’s also worth noting that the term “fake news” and its visual signifiers (a red stamp across an image, for instance) has also been appropriated more widely by websites, organisations and political figures identified as untrustworthy by fact-checkers to undermine opposing, verified reporting and news organizations” (Ward, 2017).

(3) Vaguely defined as ‘false or biased information spread through the media to influence the public debate’, fake news finishes to be equated to previous forms of misinformation and in particular to propaganda, which has been a major topic of research since the beginning of the 20th century (cf. among many others, Lasswell, 1927, Ellul, 1965 and Chomsky, 1991). Of course, fake news is said to be specifically related to digital platforms and social media (rather than traditional broadcast), yet the attributes generally used to define it are the one usually associated to propaganda – e.g. the fact of containing deceitful or inaccurate content; of pushing specific political agendas; of aiming at emotional reactions rather than rational consideration; of being echoed by different sources; of being amplified by word-of-mouth. This suggests that some of the features of fake news may not be as new as they are presented, but it also suggests that we may be missing the specificity of the phenomenon. Paradoxically, alarms for the *unprecedented* rise of fake news tend to mobilise the same concerns that have long been raised against classic propaganda, instead of focussing on the specific dangers of this contemporary form of misinformation (Jack, 2017).

(4) Presenting the advent of ‘fake news’ as the beginning of a new ‘post-truth era’ (Keyes, 2004) is also misleading because it presupposes that there was a time in which the distinction between true and false was unproblematic. Now if there a lesson to be learned from half a century of *Science and Technology Studies* (the discipline where I come from) is that the separation between true and false has never been straightforward (Lynch, 2017). To be sure, this does not mean that true and false are the same, but it does mean that their opposition is not binary or timeless. As STS scholars have repeatedly shown, a Manichean distinction between true and false is not enough to capture the vast spectrum of reliable-yet-not-without-uncertainties status of scientific theories. Even more important: the true/false dichotomy fails to render the way in which enunciates are solidified through the work of scientific laboratories (Latour & Woolgar, 1979). Far from being established by sheer force of evidence, scientific facts are built and upheld by a complex and patient work of ‘truth-grounding’ – and this applies *a fortiori* to journalism, whose truth-grounding procedures are less strictly formalised. The notion of ‘fake news’ is misleading because it supposes that malicious pieces of news are manufactured, while truthful

ones correspond directly to reality. This denies the very essence of journalistic mediation, the work to select, combine, translate and present different pieces of information in a news story. The distinction worth making is not between manufactured and unaffected information, but between well-crafted stories and badly fabricated ones.

(5) Finally, ‘fake news’ conveys the idea that the main purpose of these stories is to trick readers into believing them. While this is sometimes the case, many stories labelled as fake news circulate without asking the ‘cognitive adherence’ of those who spread them. Some fake stories are published in satirical pages that do not hide their untruthfulness; others are put forward by news outlets that play out front their ideological biases; others are just titles used to lure readers into clicking on banners or opening pages. When it comes to fake news, truth is often overrated:

while fact-checkers belonging to what elsewhere has been called the “reality-based community” (Mankoff, 2016) would define fake news as false, misleading or else unverifiable information packaged as news, fake news might do entirely different work for users who share it on Facebook. This might include acting as monetisable clickbait for viral content pages, doing issue work for grassroots activist groups, grassroots campaigning work for political loyalists and providing humour for entertainment groups (Bounegru et al, forthcoming).

Eat the cake and critique it too

While the five critiques discussed above may not be exhaustive, they are enough to discredit ‘fake news’ as a productive scientific notion – and enough to embarrass me for my supposed expertise of it. But there is worse. If these *negative* arguments (these very good claims about what fake news is not) make me feel uncomfortable, there is a *positive* one that makes me feel downright guilty. Because, to be sure, it is possible to talk *positively* about what fake news. It is not because something cannot be precisely defined that its effects cannot be observed and it is not because something has antecedents that its elements of novelty cannot be acknowledged. If the expression “fake news” has become so popular, if dozens of academic projects have been started on the subject, if hundreds of seminars have been organised, if thousands of newspaper articles have treated the question, it cannot only be because of a vast misunderstanding or an effect of fashion. Or rather, it is precisely this effect that is most interesting.

In my experience, I first stumbled on this ‘fashion effect’ while following the 2017 French presidential election. In order to set up a ‘methodological recipe’ to study fake news circulation of (see Bounegru et al, 2017), Mathieu Jacomy, Anders Munk and I decided to follow the story that most resembled to the archetype of fake news – the French equivalent of the ‘Pope endorses Trump’ hoax (www.snopes.com/pope-francis-donald-trump-endorsement/). We settle for a story about Emmanuel Macron (who ended up winning the election) being homosexual and supported by a gay lobby. The most interesting thing we discovered following this story was that its truth was never the question. While dozens of websites and social media accounts retransmitted the story, the vast majority of them did it by explicitly labelling it as fake news. Apart from the original publication on the Russian information agency Sputnik News (fr.sputniknews.com) and from a few very marginal Twitter accounts, few sources credited the story. Most venues cited the story to debunk it and most importantly to exhibit the trophy of a French fake news.

Mindful of the example of the 2016 US presidential campaign, many observers were expecting the French election to become the theatre of a similar proliferation of misinformation. For several reasons (one of which I will discuss below), this was not the case and commentators had few examples to chew on. The best one was the 'Macron is gay' one, not because anyone ever believed it (or cared about it), but because the story beautifully incarnated the fake news imagery: it involved the Russian propaganda; had sexual implications; resonated with rumours about Macron's wedding, etc. While the 'Macron is gay' story gained little traction, the 'Russian propaganda helps French trolls to spread slanders about Macron' story was a success: eventually, France had its own fake news to be shocked about. Soon even Sputnik News started to publish meta-article discussing the circulation of the story rather than its content.

The example suggests that 'virality' – rather than 'fakeness' – should be considered the birthmark of fake news. Much more than to classic propaganda, fake news resembles to Internet memes (Shifman, 2013). It does not spread because people believe in it (though this may sometimes be the case); it spreads because people (journalists, bloggers, social media influencers and social media users) like to talk about it. Fake news is the news equivalent of 'LOL cats'. Its contagiousness is such that it also spreads to most of the initiatives countering them.

I can't help but laugh at the irony of folks screaming up and down about fake news and pointing to the story about how the Pope backs Trump. The reason so many progressives know this story is because it was spread wildly among liberal circles who were citing it as appalling and fake. From what I can gather, it seems as though liberals were far more likely to spread this story than conservatives. What more could you want if you ran a fake news site whose goal was to make money by getting people to spread misinformation? Getting doubters to click on clickbait is far more profitable than getting believers because they're far more likely to spread the content in an effort to dispel the content. Win! (Boyd, 2017a)

Fake news debunking is like an autoimmune disease: it starts as a healthy response against a viral infection, but risks causing an exaggerated mobilisation becoming almost as harmful as the infection itself. Fake news is toxic not because it promotes the belief in false information, but because it saturates the public debate (and to some extent the academic debate) with a deluge of repetitions and variations of the same memes. Fake news pollutes debate like an algal bloom – its proliferation fill up the systems drawing oxygen all other discussions. This reduces the diversity of the public debate and prevent important stories to be heard and discussed.

So, demystifying efforts end up amplifying the problem that they are supposed to solve and, very importantly, this applies both to debunking of specific fake stories and criticizing the notion of 'fake news' (as I did in the first part of this paper). Far from redeeming fake news scholars, assuming a critical attitude only makes things worse. Because, by discrediting fake news, we are still talking about it. While it is easy to make fun of debunking initiatives and to demonstrate that fact-checking is counterproductive, isn't this committing the same sin that we are lecturing against? When studying a popular subject, it is too easy to 'go critic' – too easy to get the visibility that comes from fashion, but not the shame of being fashionable.

Gabriel Tarde's dream and nightmare

Critiquing the notion of 'fake news', to be sure, is better than hyping it, but it is not enough. If we are complicit in drumming the noise, we should at least try to make some good music out of it.

And this is exactly what the *Public Data Lab* (or PDL) is trying to do. The PDL (publicdatalab.org) is a network of European researchers working on digital data and digital methods. Its distinctive feature is an approach to media studies that is both pragmatic and interventionist. *Pragmatic*, because rather than studying the nature of digital mediation in theory, we are interested in using the traces or inscriptions produced through this mediation (Venturini et al., 2017) to inquire social dynamics in practice. *Interventionist*, because this inquiry is (in a distinctively deweyan way – Dewey, 1938 and especially 1946) not only the observation of an independent state of affairs, but also the deliberate engagement with a social issue with the deliberate agenda of promoting an open debate around it. This does not mean that we always intend to push things in one direction or another (though we sometimes might), but that we strongly hope that our interventions will not leave unaffected the subjects we study.

When we started to work on fake news, we had these two ideas clearly in view. We wanted to draw from its contagiousness and turn it from a curse into an occasion. Two occasions in fact: an occasion to rethink media studies in more *pragmatic* way and an occasion to *intervene* in the debate on the organisation of the media system.

To seize the first occasion, we launched the *Fake News Field Guide* with the objective of displacing the discussion from what fake news is or why it is dangerous to the question of how it can be empirically investigated. That is why the Field Guide is a collection of methodological ‘recipes’. By playing with the cookbook genre, we tried to tone down the fake news drama and suggest that the study of fake news opens interesting perspectives on contemporary media systems. We identified five of such perspectives (but the list is certainly not exhaustive): investigating ‘fake stories’ to understand how public Facebook pages create different publics and discussion hotspots (chapter 1); following ‘fake stories’ to trace how viral information circulates by crediting and discrediting actions (chapter 2); looking at the trackers employed by different news websites to trace and monetise their audience (chapter 3); investigating how ‘fake stories’ are circulated not only in the form of classic news article, but also as memetic images (chapter 4, forthcoming); studying how ‘fake stories’ are used in the context of online trolling campaigns (chapter 5, forthcoming).

From the list above, it should be clear that we were less interested in fake news itself than in a series of more general dynamics of online media. Because of their relative simplicity (fake stories are usually flatter than ordinary news story), their exaggeration (which makes it easier to detect and follow them), their geographical spread (scholars have identified examples of it in most countries of the world), their rapid reproduction (fake news tends to rise and fall in a few weeks and often of in a few days), their elevated mutation rate (I will expand this argument below), fake news may be the *drosophila melanogaster* of media studies. The drosophila is a little and relatively common fly (often found around ripe fruits), which assumed a crucial importance in the history of genetics. Because of reasons similar to the ones just listed for fake news, the drosophila became the ‘model organism’ for genetic research allowing to qualify and quantify a series of intuitions about gene reproduction (Kohler, 1994). As the drosophila facilitated the birth of experimental chromosome mapping, so fake news may help the development of new forms of *empirical* media mapping.

Readers will have noticed that none of the reasons why fake stories are interesting research subjects have anything to do with their falseness. I said above that French presidential campaign

was characterized by a relatively low level of misinformation. One of the reasons may be that in the months preceding the elections, the French debate was occupied by the so-called ‘affaire Fillon’. On January 25th, the satirical weekly *Le Canard Enchaîné* published a piece alleging that Penelope Fillon, wife of François Fillon (candidate of the Republican party and strong favourite for the election), had been paid for years for a fictitious work of parliamentary assistant. The scandal colonized traditional and social media until the elections in April and, according to many, caused Fillon to be excluded at the first round of the election.

While the Fillon affair has nothing of a ‘fake news’ story (and was instead based on a solid journalistic investigation), it ended up spreading like one and occupying the public debate in a similar way (thereby filling the space needed by fake news to develop).

As the expression goes, the Fillon affair *went viral*. This expression deserves to be questioned. While the metaphor of viral epidemics is commonly used to characterize fake news (I have done it myself just earlier), the similitude is often used for the wrong reasons. As Jenkins *et al.* (2013) rightly pointed out, this metaphor tends to present digital publics as passive recipients of an infection that they cannot but pass along:

“the viral metaphor does little to describe situations in which people actively assess a media text, deciding whom to share it with and how to pass it along. People make many active decisions when spreading media, whether simply passing content to their social network, making a word-of-mouth recommendation, or posting a mash-up video to YouTube. Meanwhile, active audiences have shown a remarkable ability to circulate advertising slogans and jingles against their originating companies or to hijack popular stories to express profoundly different interpretations from those of their authors” (Jenkins *et al.*, 2013, p. 20).

The virality metaphor can be saved, however, by reminding that actual viruses are dangerous not only because they are contagious, but also because there are capable of mutating. While spreading from one organism to the other, biological viruses are transformed by their interaction with other viruses and with their hosts and vectors of diffusion (Sanjuán & Domingo-Calap, 2016). And the same happens to viral news, which is not just passed along by news outlets and Internet users, but actively modified by them. It is through this mutation that viral news (but also image memes, online jokes, buzz advertisements, etc.) can saturate public debate. A news story (fake or true) can acquire a bursting visibility if it sufficiently eye-catching, but it will not thrive unless it is altered and retransmitted in a swarm of different variants.

This is exactly what happened in the case of the Fillon affair. If the news had stopped at the first article by *Le Canard Enchaîné* and its reprises by other media, the story would not have occupied the French public debate for almost three months. If it did, it is because the first allegation was followed by a proliferation of other accusations concerning different underserved wages of Fillon’s wife; different members of the Fillon family; and other questionable conducts of François Fillon. Each of these accusations, and particularly those concerning Penelope Fillon, produced a vast range of reactions, denials, clarifications, comments, glosses, critiques, jokes, etc. which all contributed to maintaining the affair at the centre of French electoral debate.

Not unlike rumours (Morin, 1969), digital viral contents proliferate by transmission and transformation. In this, they provide a perfect illustration of the mechanism through which social phenomena are constructed according to Gabriel Tarde (1890). In his famous dispute with

Emile Durkheim over the fundaments of the nascent sociology, Tarde refused the idea that underlying or emergent structure was at the basis of collective phenomena and claimed instead that their existence was to be searched in the ‘simple’ imitation of individual behaviours as well as in the progressive alteration that it entails (Latour, 2002). Tarde, however, found it difficult to defend its position *empirically*, because the research methods available at the time did not allow to follow the transmission and transformation of collective actions at the scale and with the sharpness demanded by his argument. This may be possible today thanks to the capacity of digital technologies to store and retrieve each of these movements of diffusion and mutation (Latour *et al.*, 2012 and Boullier, 2015).

Fake news is both the dream and the nightmare of Tarde’s sociology. The dream, because it offers an opportunity to trace the transmission and transformation of collective actions and thereby identify the different ‘regimes’ of circulation and mutation occurring in the contemporary media systems; the nightmare, because it represents the pathological degeneration of such mechanism.

Pès el tacòn del buso

But there is more. The study of fake news is not only a methodological occasion for the social sciences, but also and crucially an occasion to *intervene* in the debate on the organisation of the media system. As scholars have not failed to notice, fake news represents the dark side of a quality for which digital media have long been celebrated. As well known, the Internet has a double filiation. On the one hand, the idea of a distributed computation network derived from the need to protect military intelligence from nuclear attacks and found its early sponsor in the Defense Advanced Research Projects Agency (DARPA). On the other hand, Internet technologies were developed by circles of fringe engineers inspired by American counter-culture (Cardon, 2010). In these communities, the decentralized nature of the new telecommunication protocols (granting, in principle, the same importance to all nodes of the network) was strongly associated with egalitarian utopias.

“The hacker culture clearly had certain points in common with the hippie counterculture and with Arpanauts’ representations. It shared the same refusal of centralized and commercial information technology that IBM symbolized at the time. The main difference between the two cultures lay in hackers’ far broader view of the use and future of IT. For them it was not only an intellectual tool for academics but also a device to put into everyone’s hands, capable of building not only new invisible colleges but also a new society” (Flichy, 2007 p. 67).

The Internet, indeed, offers something that no media has ever offered before: a ‘many-to-many’ communication. Before its advent, communication could either be bidirectional but limited to few people (as in face-to-face encounters, letters and telephone calls) or open to many people but unidirectional (as in megaphones, newspapers, radio and television). Digital media makes it technically possible for many people to hold bidirectional discussions with many others, a situation that was promptly associated with that of the mythic Greek *agoras*. In a speech to the International Telecom Union (Buenos Aires, 21 March 1994), for example, Al Gore (at the time US vice-president) emphatically announced:

“In a sense, the Global Information Infrastructure will be a metaphor for democracy itself... it will in fact promote the functioning of democracy by greatly enhancing the participation of

citizens in decision-making. And it will greatly promote the ability of nations to cooperate with each other. I see a new Athenian Age of democracy forged in the fora the Global Information Infrastructure will create."

It is important to notice that the democratic promises of digital media are here directly connected to their capacity to remove communication asymmetries, bypass conventional gatekeepers and offer each citizen the possibility to become an independent source of information. Sadly, but not entirely surprisingly, these are the very same features that create the abnormal news virality we are mourning these days. If fake stories spread and mutate cancerously, it is to a large extent because digital media have spectacularly increased the number of potentially contagious interactions.

In its most basic version, this argument is simply a tribute to traditional journalistic institutions. For almost two centuries, since the application of the printing press to the reporting of news, journalism has developed both as an ideology (Deuze, 2005) and as an organisation (Pettegree, 2014). With its increasing professionalization, journalism has also obtained (at least in western countries) a growing monopoly in the distribution of news and, while not without shortcomings and dysfunctions, it has contributed to maintaining a (more or less) functioning public debate, according to the old lippmannian lesson (Lippman, 1922).

Yet, with the advent of digital media, professional journalism has entered a structural crisis. By multiplying the sources of information and entertainment, electronic technologies have multiplied the competing sources of information and entertainment and undermined both its traditional sources of financing (advertising and circulation). According to the Pew Center 2015 *State of the US News Media* report (www.journalism.org), the revenue of newspapers has constantly decreased in the last decades, falling from almost 50 billion dollars in 2005 to less than 20 in 2014. Television news has done slightly better, but its revenue is also stagnating. The crisis is particularly harsh for local news, with dozens of local newspapers closing their doors (cf. www.newspaperdeathwatch.com). To be sure, new forms of reporting are sprouting up on digital media. However, whether these experiments will be enough to support a constant and professional coverage of social debates with adequate human and technical resources is still uncertain. For the moment, the number of journalists employed in US newsrooms has dropped from around 55 thousand in the early 80's to below 37 thousand in 2013 (again according to the Pew Center 2015 report).

The crisis of professional journalism, however, is only one aspect of a deeper transformation of contemporary media systems. The issue here is not simply that traditional newsrooms are suffering the competitions of new information outlets, offering the same 'product' through different distribution channels. The issue is that the very notion of 'news' is fading in the new media environment. Unlike what early observers believed, the many-to-many communication introduced by digital media does not resemble the older forms of broadcast (not only at least). This idea was plausible in the first years of the Internet, when every website functioned as a little broadcasting station, but has been shattered by the advent of the so-called social media platforms. Blurring the distinction between narrowcast and broadcast communication, Internet platforms have also eroded the distinction between private and public life.

Consider Facebook, for instance. In principle, the platform distinguishes between private profiles and public pages (reserved to celebrities, brands, institutions and open groups). In

practice, however, the interface and underlying functioning of the two types of “news feeds” are perfectly equivalent. The only difference is that while users can restrict the access to the contents on their profile (selectively choosing what to share with whom), pages are open to everyone. Yet, when you follow a page or befriend a profile, the way in which their posts are channelled in your news feed can be hardly distinguished: your annoying friend from high school and the New York Time speak to you in the exact same way.

Again, the problem is not just that public news and private messages compete for citizens' limited attention (this has always been the case, as Walter Lippmann noted in *The Phantom Public*, 1927). The problem is that the disappearance of the frontier between the private and public sphere challenges the distinction between private and public law that regulates collective life since the rise of the modern nation-states (in western democracies at least). Going back to the Facebook example, when studying the circulation of fake news on the platform, we were constantly frustrated by the impossibility to follow viral contents in all their movements. While users and contents travel seamlessly through the network of news feeds, scholars are blocked at nonexistent frontier of private profiles. This is, of course, reasonable and even necessary to protect the privacy of Facebook users, but does preclude the study news mutation and contagion.

A cross-cultural comparison may help put the problem in focus. In a recent study Gary King, Jennifer Pan and Margaret E. Roberts (2017) investigate the strategy of social media influence of the Chinese government (the so-called “50 Cent Party”) and conclude:

“We estimate that the government fabricates and posts about 448 million social media comments a year. In contrast to prior claims, we show that the Chinese regime’s strategy is to avoid arguing with skeptics of the party and the government, and to not even discuss controversial issues. We show that the goal of this massive secretive operation is instead to distract the public and change the subject” (p. 484).

What is most interesting is that the Chinese government decided to respond publicly to the paper in an editorial published in a state-owned newspaper (<http://opinion.huanqiu.com/editorial/2016-05/8958840.html>). In its response, the government admits its intervention in online discussions and in fact take credit for it:

“It is obvious that the team doing this research has only the most superficial knowledge of China’s national circumstances, muddling the distinctions between official authoritative information, the official news media, and ordinary online statements; they also fail to understand the legitimacy of “public opinion guidance” within the Chinese system. They take the structures and mechanisms of the Western media field as the standard ...

The Chinese internet media’s largest problem is not being dominated by the “Fifty Cent Party,” but rather the amplification of negative and alternative information on Chinese domestic issues caused by opinion formation mechanisms that have been a part of the Internet since it was invented in the US; Chinese society, in the midst of a transformation, does not have the hedging mechanisms to deal with this amplification, so traditional public opinion guidance systems don’t seem to be pulling their weight when it comes to overcoming these problems. The Internet media space has an infinite capacity but its borders and its core are unclear, so some grassroots social issues are always able to suddenly attract the attention of the entire Internet, creating one hot button issue after another in the online Chinese media” (Appendix B of King, Pan & Margaret 2017).

One can, of course, disagree with the conclusion but the argument is not unfunded. With shocking candour, the Chinese government justifies its intervention by pointing precisely at the blurring of the private and public spheres: if state guidance is desirable and even necessary, they claim, it is because social media amplify private and grassroots issues and allow them to rise to the core of public opinion.

The Chinese strategy provides another example of a cure that turns out to be worse than the disease – or, as in a beautiful saying from the Italian city of Padova, “xe pèsò el tacòn del buso” (the patch is worse than the hole). If fake news cannot be solved by fact-checking initiatives, they cannot be solved by an increase of central control either. Such solution is incompatible with democratic debate when it is implemented by a state, and even more when implemented by a media themselves. It is a very bad idea to ask Facebook, Google, YouTube or any other platform to watch over public debate and it is illusionary to believe that algorithms can be developed to identify and eliminate fake news. Given the impossibility to demarcate fake news, such algorithmic solutions will be at best useless and at worse hide censorship initiatives under the fake premise of mechanical objectivity.

As I tried to show, the problem with fake news comes from their virality and their virality comes from their capacity to exploit the multilateral communication of digital media to spread and mutate. Having its roots in the vanishing separation between public and private communication, the problem runs much deeper than we tend to believe. It is our duty as communication scholars to denounce the reductionist solutions that create more problems than they solve, but also to contribute opening a serious debate about the private and public in the contemporary media systems.

References

- Boullier, D. (2015). Socio Vie et mort des sciences sociales avec le big data. *Socio*. 4. pp. 19–37.
- Bounegru, L., Abildgaard, M.S., Birkbak, A., Gray, J., Jacomy, M., Elgaard, T.J., Koed Madsen, A. & Munk, A.K. (2017). Main Document – with full author details. *Forthcoming*.
- Bounegru, L., Gray, J., Venturini, T. & Mauri, M. (2017). *A Field Guide to Fake News: A Collection of Recipes for Those Who Love to Cook with Digital Methods*. [Online]. Available from: <https://ssrn.com/abstract=3024202>.
- Boyd, D. (2017). *Did Media Literacy Backfire?* [Online]. 2017. Points. Available from: <https://points.datasociety.net/did-media-literacy-backfire-7418c084d88d>. [Accessed: 8 September 2017].
- Cardon, D. (2010). *La démocratie sur Internet. Promesses et limites*. Paris: Seuil.
- Chomsky, N. (1991). *Media control : the spectacular achievements of propaganda*. New York: Seven Stories.
- Claire, W. & Derakhshan, H. (2017). *Information Disorder: Toward an interdisciplinary framework for research and policymaking (Report to the Council of Europe)*.
- Deuze, M. (2005). What is journalism? *Journalism: Theory, Practice & Criticism*. [Online]. 6 (4). pp. 442–464. Available from: <http://journals.sagepub.com/doi/10.1177/1464884905056815>.
- Dewey, J. (1938). *Logic: The Theory of Inquiry*. New York: Holt, Rinehart And Winston.
- Dewey, J. (1946). *The public and its problems: an essay in political inquiry*. [Online]. Chicago: Gateway Books.
- Ellul, J. (1965). *Propaganda: the Formation of Men's Attitude*. New York: Random House.
- Flichy, P. (2007). *The Internet Imaginaire*. Cambridge Mass.: MIT Press.

- Jenkins, H., Ford, S. & Green, J.B. (2013). Spreadable media. *Postmillennial pop*. [Online]. p.p. 351. Available from: <http://gso.gbv.de/DB=2.1/PPNSET?PPN=804244480>.
- Keyes, R. (2004). *The Post-Truth Era: Dishonesty and Deception in Contemporary Life*. New York: St. Martin's.
- King, G., Pan, J. & Roberts, M.E. (2017). How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, Not Engaged Argument. *American Political Science Review*. [Online]. 111 (3). pp. 484–501. Available from: https://www.cambridge.org/core/product/identifier/S0003055417000144/type/journal_article.
- Kohler, R.E. (1994). *Lords of the Fly. Drosophila Genetics and the Experimental Life*.
- Lasswell, H.D. (1927). *Propaganda Technique in World War*. Chicago: Chicago University Press.
- Latour, B. (2002). Gabriel Tarde and the End of the Social. In: P. Joyce (ed.). *The Social in Question. New Bearings in the History and the Social Sciences*. London: Routledge, pp. 117–132.
- Latour, B., Jensen, P., Venturini, T., Grauwin, S. & Boullier, D. (2012). 'The whole is always smaller than its parts': a digital test of Gabriel Tardes' monads. *The British Journal of Sociology*. [Online]. 63 (4). pp. 590–615. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23240834>.
- Latour, B. & Woolgar, S. (1979). *Laboratory Life. The Construction of Scientific Facts*. Los Angeles: Sage.
- Lippmann, W. (1922). *Public Opinion*.
- Lippmann, W. (1927). *The Phantom Public*. New York: The Macmillan Company.
- Lynch, M. (2017). *Post-truth, alt-facts, and asymmetric controversies*. 2017. First 100 Days.
- Morin, E. (1969). *La rumeur d'Orléans*. Paris: Seuil.
- Pettigree, A. (2014). *The Invention of News: How the World Came to Know About Itself*. New Haven: Yale University Press.
- Sanjuán, R. & Domingo-Calap, P. (2016). Mechanisms of viral mutation. *Cellular and Molecular Life Sciences*. 73 (23). p.pp. 4433–4448.
- Shifman, L. (2013). *Memes in Digital Culture*. Cambridge Mass.: MIT Press.
- Tandoc, E.C., Lim, Z.W. & Ling, R. (2017). Defining 'Fake News'. *Digital Journalism*. [Online]. 811 (August). pp. 1–17. Available from: <https://www.tandfonline.com/doi/full/10.1080/21670811.2017.1360143>.
- Tarde, G. (1890). *Les lois de l'imitation*.
- Venturini, T., Jacomy, M., Meunier, A. & Latour, B. (2017). An unexpected journey: A few lessons from sciences Po médialab's experience. *Big Data & Society* 4 (2). Available from: <http://journals.sagepub.com/doi/10.1177/2053951717720949>.
- Zuckerman, E. (2017). *Stop saying 'fake news'. It's not helping*. 2017. <http://www.ethanzuckerman.com/blog/2017/01/30/stop-saying-fake-news-its-not-helping/>.

Search chicagoreader.com

GO



READ | SHARED | COMMENTS

Like 0

Tweet

G+

Share

< Dead Writers Collective now dead af...

| The One-Minute Festival offers a ta... »

WEDNESDAY, FEBRUARY 22, 2017

NEWS / MEDIA

Buzzfeed media editor Craig Silverman on the impact of fake news, what's next for it, and what, exactly, it even is

Posted By Austin Brown on 02.22.17 at 02:07 PM



Craig Silverman at the University of Chicago's International House

UNIVERSITY OF CHICAGO INSTITUTE OF POLITICS

Craig Silverman, media editor of Buzzfeed, was showing a picture of an ABC News article recently to a crowd at the University of Chicago's International House Assembly Hall. The headline read "Donald Trump Protester Speaks Out: 'I Was Paid \$3,500 To Protest Trump's Rally.'" It looked official enough. That is, until the audience began to notice things like the ABC logo font in Arial, the odd kerning, the fact that paid protesting really *isn't a thing*, and finally,

BLEADER

Joseph Chilliams talks about the losses he overcame to release one of the year's best Chicago hip-hop albums

Pivot Gang cofounder Joseph Chilliams released his debut, *Henry Church*, in August, after months of career highs and personal lows—including the death of friend and collaborator Dinner With John.

By Leor Galil | 09.27.17

BLEADER

Judge in Laquan McDonald murder case faces a challenge trying to control the flow of information, and other Chicago news

Also, Rahm Emanuel says Trump is trying to distract Americans from important issues like the North Korea crisis.

By Kate Shepherd | 09.28.17

BLEADER

The LatinxArts festival gets on its feet with a show of dances by immigrant choreographers

Michel Rodriguez Cintra of Lucky Plush, Wilfredo Rivera of Cerqua Rivera Dance Theatre, and Victor Alexander of Hedwig Dances represent in Movements, at Hairpin Arts Center Friday.

By Oliver Sava | 09.27.17

BLEADER

The mythical idea of the American heartland shouldn't define the midwest ↗

The heartland is a fictional notion of a midwest without cities, black people, and immigrants.

By Daniel Kay Hertz | 02.15.16

PRESENTED BY SAAMBAA

See Top Trending Food & Fun

BURLINGTON

~~Les fausses nouvelles : le nouveau visage d'un vieux problème~~
the author's name—*Jimmy Rustling*, a nod to the popular 4chan meme, "That Really Rustled My Jimmies."

13

powered by **saambaa**

This is "fake news"—not in the way President Donald Trump uses the term, which is basically to describe any story that seems unflattering, or the meme it's become, which is anything I disagree with, but as it was originally intended—purposefully invented or falsified news (especially politics) stories, written for low-budget websites and dramatized to receive maximum click-through traffic.

Silverman was invited by the U. of C.'s Institute of Politics to give a talk, called "Fake News, Alternative Facts, and the World of Misinformation," on February 13—and who better? Silverman's made his name fact-checking and managing misinformation, first as the "Regret the Error" fact-checking and media verification columnist for Poynter, then as the founder of Emergent.info, a start-up meant to tend to the ever-growing rumor mill in online news. At Buzzfeed, he spearheaded much of the site's forward-thinking coverage of online misinformation outlets from Macedonia to Canada. With "fake news" on the lips of every politician and editor, it would seem this is Silverman's time to shine.

But at the beginning of the last week's talk, he demurred. "I think the term 'fake news' has almost been rendered meaningless at this point," he said. "It means a lot of different things to different people." Instead, he simply hoped to set the record straight, to "offer a little clarity, at least in terms of how I see fake news and how I define it, but also what I call the 'online misinformation ecosystem' as well." It would go on like this for the remainder of the talk—Silverman delving into his wealth of knowledge on how we get tripped up by falsehoods on the Internet, but always leaving questions open as to where the stories will go next.

The fake story about a paid protester that Silverman led off with might have just languished in Facebook and Twitter feeds far from the gaze of journalists and the political class if not for two people—Corey Lewandowski and Kellyanne Conway, seeming to take the piece as fact, had both tweeted the story out to their huge numbers of followers. It's a narrative that's become familiar at this point, and Silverman stressed the real concern: "This is really important for misinformation," he said. "When people in positions of power and authority share it, and put it out there, it adds to its perception of credibility."

These highly charged political stories have become an all-too-regular occurrence—a person with little credibility comes up with either a misrepresented statistic or a blatantly false idea, and the Trump administration runs with it, terrifying many on the left while energizing those in the president's base, many of whom already believe they've been lied to.

Silverman emphasized that while many of those sharing fake news are politically motivated, the stories themselves usually have less partisan origins. Many of the most popular stories, he made clear, were from "trolls who don't necessarily have an ideological motive, who are just out there causing trouble."

Silverman made sure to emphasize that while there's great potential for fake news websites (which he's also compiled an ever-growing list of) to be used by political actors to set an agenda, the sites that he's found are instead driven by the simple desire to make a quick buck.

True fake news today, Silverman explained, "has to be 100 percent false. It has to be purposefully created as false, and it has to be financially motivated." It's a "cynical form of attention harvesting" that takes advantage of people's reliance on social media for emotional validation just as much as for information gathering, hijacking the distrust of mainstream media sources and the endless information that the Web can provide.

So these oft-cited Macedonian teens running some of the biggest fake news sites might be associated with Russia, sure. But it's telling that in Silverman's own investigation, a 17-year old site owner say nothing about political motivations. "In Macedonia the economy is very weak and teenagers are not allowed to work, so we need to find creative ways to make some money," the teen says. "I'm a musician but I can't afford music gear. Here in Macedonia the revenue from a small site is enough to afford many things."

Of course, Silverman went on to remind people, fake news didn't come out of nowhere. Misinformation, intentional or otherwise, is as old as dirt.

At one point, Silverman brought up the *Sun*, a now-defunct New York City broadsheet that at one point coexisted with the *New York Times* in the mid-1800s. An early benefactor of the "penny press," which made it cheaper to print papers and thus led to the rise of mass media, the *Sun* also took advantage of another trick that Silverman described as "just print crazy shit." This included a six-part series about an explorer who had seen what was on the moon . . . and found it to be filled with a species that was half man and half bat.

But while fake news isn't new, the combined set of circumstances surrounding its recent flourishing is indeed novel: partisan drama, a networked society, and an extension of the endless craving for attention-grabbing stories first built up by the 24-hour cable news cycle. Plus, who needs a penny press when you can just buy a domain name off GoDaddy?

The election, in particular, Silverman said, "combined all the things that will make misinformation run rampant. One, a presidential cycle attracts a huge amount of attention . . . and this was an especially crazy cycle, so even more attention. That's good for attention harvesters. That's what they need."

Silverman went on: "The emotion and strong-beliefs component [are] really, really important, 'cause we know people react to stuff that reinforces what they want to hear, things that get an emotional reaction. Reading something that makes you happy doesn't make you react as strongly as something that makes you angry or disgusted, and people know that."

"And the other thing is," Silverman said, "in general in politics there's a good amount of spin and misinformation. Donald Trump is, frankly, very unique because of the amount of totally false things that he says. And so when you have a major candidate who will consistently say things that are not true, it sort of lays the groundwork and opens the area for lots of other crazy things that are not true to be said as well."

"The last big one is the social network platforms and algorithmic filtering. I don't think Facebook had any idea that a lot of stuff that was false and misleading was getting as much traction as it was—their platform is so big it's impossible to monitor it at that level. Same with Google."

This perfect storm of factors came together, and as Election Day approached, Silverman showed that the top 20 fake news stories had received significantly more engagement on Facebook than the top 20 from mainstream news outlets. The revelation reflects the potentially problematic conflations that have seeped in through Facebook and Twitter's catch-all interfaces—between professional and personal, emotional and intellectual, political and economic.

With the election in the rearview, one might be forgiven for thinking that fake news might similarly slide out of view, a relic of a tumultuous time that seems absurd, even with respect to the Trump administration's chaotic first month.

But Silverman's recent reporting bears out differently. He recently published a story noting that the film *A Cure for Wellness* was using fake news as part of a marketing campaign, in a sort of twisted meld of alternate-reality game and the misinformation economy.

And that may be the strangest implication of all. Fake news might have had its origins as a kind of DIY online hucksterism, a peripheral to trolling that has its roots in the same ethos—but after

Les fausses nouvelles : le nouveau visage d'un vieux problème.

Tags: fake news, Buzzfeed, Craig Silverman, Facebook, University of Chicago, U. of C., Twitter, Image

Make a difference on topics related to this story. Powered by Public Good.



Take Action

Like 0

Tweet

G+

Share

RELATED STORIES



Is it time to retire the concept of 'fake news'?

Let's instead try "complete bullshit" and other terms for describing lies dressed up as the truth.

by Michael Miner

TODAY



NEXT EVENT



Blackhawks at Red
Wings
7:30 PM

PRESENTED BY STANZA



TODAY

White Sox vs. Angels
8:10 PM

TODAY

White Sox vs. Angels
8:10 PM

WATCH

MORE BY AUSTIN BROWN



MUNA find compassion and hope in conflict

by Austin Brown



Khalid's small-scale groundedness sets him apart from the contemporary R&B pack

by Austin Brown



Jayda G is dedicated to creating a dance party for the freaks

by Austin Brown

MORE >

Social media manipulation by state actors: fake news, political bots, and patriotic trolls

Samantha Bradshaw

D.Phil. Candidate, Oxford Internet Institute

Researcher, Computational Propaganda Research Project, Oxford University

Senior Fellow, Canadian International Council

Presentation Summary

Social media has become a critical platform for politics. Individuals around the world use social media to find important news and information that they use to develop their political values and identity. Increasingly, state and non-state actors have co-opted social media technologies to spread disinformation and computational propaganda during critical political events. Throughout the elections of 2016, lies conspiracy theories and alternate realities flowed freely across the media ecosystem. In the United States, some of the most widely shared stories about politics were fake news stories. In Michigan, a key battleground state, fake news was shared just as widely as professionally produced news, with “junk news” stories outperforming high-quality news the day before the election (Howard et al 2017). Similar amounts of junk news spread during the elections in France, Germany and the United Kingdom, with the amount of junk being shared on social media slowly increasing as election days drew nearer (see Desigaud et al 2017, Kaminska et al 2017, Neudert et al 2017).

Coordinated efforts to spread junk news over the Internet have contributed to the spread of cynicism and increased the division between citizens, parties, and political ideologies. Although fake news is often discussed in the context of the 2016 US Presidential election, the spread of disinformation over the Internet and social media is a global phenomenon that is threatening democracies around the world. A number of states have skilfully combined information operations with electoral espionage to further pollute political discourse and destabilize democratic institutions (Bradshaw and Howard, 2017). Computational propaganda is used by authoritarian regimes as a tool of social control. Often, political bots are deployed around a particular issue, election, or personality, in order to give a false sense of popularity, momentum, or relevance to fringe or minority dispositions. In other cases, patriotic trolls target individuals who hold different or dissenting opinions, creating a chilling effect on freedom of speech across the digital public sphere. In democracies, there is evidence of political parties and other non-state actors using these same techniques to spread falsehoods and influence the broader media agenda (Bradshaw and Howard, 2017). In all instances, social media is creating new challenges for democracy.

This first part of this presentation will explore how much fake news spread in the United States, the United Kingdom, France, and Germany leading up to the elections that took place in 2016 and 2017. Drawing on the data collected and analysed by the Computational Propaganda Research Project, this half of the presentation will present data on the amount of fake news being spread on Twitter. The second half of the presentation will take a step back and look more comparatively at governments and political parties who actively spread disinformation online. It will examine the different tools, capacities, and forms these actors use to pollute the digital information environment. This presentation will conclude with a discussion of the future challenges, and the consequences of global disinformation campaigns for democracy.

References

- Bradshaw, Samantha, and Philip N. Howard. "Troops, Trolls and Troublemakers: A Global Inventory of Organized Social Media Manipulation." Working Paper. Oxford, England: Project on Computational Propaganda, July 2017.
<http://comprop.ox.ac.uk/2017/07/17/troops-trolls-and-trouble-makers-a-global-inventory-of-organized-social-media-manipulation/>
- Howard, Philip N., Gillian Bolsover, Bence Kollanyi, Samantha Bradshaw, and Lisa-Maria Neudert. "Junk News and Bots during the U.S. Election: What Were Michigan Voters Sharing Over Twitter?" COMPROP Data Memo 2017.1, March 26, 2017
<http://comprop.ox.ac.uk/wp-content/uploads/sites/89/2017/03/What-Were-Michigan-Voters-Sharing-Over-Twitter-v2.pdf>
- Desiguad, Clementine, Phillip N. Howard, Bence Kollanyi, and Samantha Bradshaw. "What Are French Voters Sharing Over Twitter Between the Two Election Rounds," 2017.
<http://comprop.ox.ac.uk/wp-content/uploads/sites/89/2017/05/What-Are-French-Voters-Sharing-Over-Twitter-Between-the-Two-Rounds-v7.pdf>
- Monica Kaminska, John D. Gallacher, Bence Kollanyi, Taha Yasseri, and Philip N. Howard. "Social Media and News Sources during the 2017 UK General Election" Data Memo 2017.6. Oxford, UK: Project on Computational Propaganda.
<http://comprop.ox.ac.uk/wp-content/uploads/sites/89/2017/06/Social-Media-and-News-Sources-during-the-2017-UK-General-Election.pdf>
- Neudert, Lisa-Maria, Philip N. Howard, and Bence Kollanyi. "What Were German Voters Sharing Over Twitter" 2017. <http://comprop.ox.ac.uk/wp-content/uploads/sites/89/2017/03/What-Were-German-Voters-Sharing-Over-Twitter-v6-1.pdf>

Stopping Fake News: A Ukrainian Example

Nadine I. KOZAK

Summary:

Responding to a state-sponsored fake news campaign, a group of volunteer Ukrainian journalist-activists turned fact checking into a counter-propaganda weapon. Our work explores StopFake.org, the organization they created in March 2014, to fight the weaponization of fake news by the Russian government, journalists, and social media content creators during the Ukraine crisis. The group employs media literacy techniques to evaluate stories for signs of falsified evidence, including manipulated or misrepresented images and quotes. This contrasts with traditional fact-checking organizations that evaluate nuanced political claims but assume accuracy in the reporting. Employing work from science and technology studies, we argue that scholars (and ideally, journalists) can acknowledge that narratives are socially constructed without having to treat all narratives as interchangeable.

This research, conducted with my colleagues Maria Haigh and Thomas Haigh, analyzes the work practices of StopFake.org, including the methods the group used to both identify and analyze news stories for potential fakes. Our work examined the 539 refutations StopFake.org published in the first 18 months of its existence, between March 2014 and August 2015. Content analysis was conducted on these published reports to categorize the debunking evidence used.

Biography:

Nadine I. Kozak is an assistant professor at the School of Information Studies at the University of Wisconsin-Milwaukee. Her research interests include information policy, regulation, and law, especially how these impact the individuals, organizations, and communities subject to them. She also explores how people interpret, adopt, respond to, and resist information law, policy, and regulation.

Detecting Deception in Online Information

by Victoria Rubin, Western University

After the 2016 Presidential Elections social media users' awareness of intentional manipulation of online content is at an all-time high, but the reliance on unverified information (often obtained from strangers) still persists. In the past several years, conceptual tools dealing with language accuracy, objectivity, factuality and fact-verification have increased in importance in various subject areas - journalism, online marketing, proofreading and politics, to name a few.

Building on years of Deception Detection research in Interpersonal Psychology, Communication Studies, and Law Enforcement, a cutting-edge technology has emerged from the fields of Natural Language Processing and Machine Learning since around 2004-2006. Spurred by demand from practitioners for stable, quick and accurate deception detection tools, scholars have begun to create software for 'deception detection'. A limited number of automated (or partially automated) online deception detection tools became available for the public by around 2010. As of late 2016, the field of automated detection has been applied to social media context relatively infrequently. By the end of 2017, there still have been only a handful of (at least well-known) works that address this problem. Nor have they received much attention in mainstream North American media coverage or in the scientific community, in spite of the extensive mainstream media coverage of fake news around the U.S. Presidential Elections.

In this talk I will introduce the audience to the area of automated deception detection research, with a cursory look at the roots of the field in pre-social media data types. I will draw attention to existing analytical methodologies and pose the question of their applicability to the context of social media. I will discuss what is currently known about people's overall abilities to spot lies and what constitutes predictive cues to tell the liars apart from truth tellers. Deception detection methods typically employ predictive modeling for measurable linguistic (content-based) features of texts such as complexity, uncertainty, non-immediacy, diversity, affect, specificity, expressiveness, and informality.

Other content verification methods for 'veracity prediction' and 'rumor debunking/busting' can further contribute towards a solution for what has been indiscriminately called 'the fake news problem' in the main stream media since fall 2016. Current rumor busters employ a multitude of user-based and network-based features such as location, account behavior, and message propagation, and are complementary to content-based features from automated deception detection.

In conclusion, social media requires content verification analysis with a combination of previously known approaches for deception detection, as well as novel techniques for debunking rumors, credibility assessment, factivity analysis and opinion mining. Hybrid approaches may include text analytics with machine learning for deception detection, network analysis for rumor debunking and should incorporate world knowledge databases to fully take advantage of the linguistic, interpersonal, and contextual awareness. These methods and resulting tools are meant to augment our human discernment, rather than replace it, by highlighting potentially false information which may require scrutiny. Social media users' critical thinking remains the key information literacy skill to navigating the increasingly toxic online environments.

[[The associated article](#) in *The SAGE Handbook of Social Media Research Methods* (2017) is accessible at
<https://us.sagepub.com/en-us/nam/the-sage-handbook-of-social-media-research-methods/book245370%20>]

Victoria L. Rubin is an Associate Professor at the Faculty of Information and Media Studies and the Director of the Language and Information Technologies Research Lab (LiT.RL) at the University of Western Ontario. She specializes in information retrieval and natural language processing techniques that enable analyses of texts to identify, extract, and organize structured knowledge. She studies complex human information behaviors that are, at least partly, expressed through language such as deception, uncertainty, credibility, and emotions. Her research on Deception Detection has been published in recent core workshops on the topic and prominent information science conferences, as well as the Journal of the Association for Information Science and Technology. Her 2015–2018 project entitled Digital Deception Detection: Identifying Deliberate Misinformation in Online News is funded by the Government of Canada Social Sciences and Humanities Research Council (SSHRC) Insight Grant. For further information, see <http://victoriarubin.fims.uwo.ca/>

Kate Starbird [Follow](#)

Asst. Professor of Human Centered Design & Engineering at UW. Researcher of crisis informatics an...

Mar 14 · 19 min read

Information Wars: A Window into the Alternative Media Ecosystem

Conspiracy Theories, Muddled Thinking, and Political Disinformation

Background: Examining “Alternative Narratives” of Crisis Events

For more than three years, my lab at the University of Washington has conducted research looking at how people spread rumors online during crisis events. We have looked at natural disasters like earthquakes and hurricanes as well as man-made events such as mass shootings and terrorist attacks. Due to the public availability of data, we focused primarily on Twitter—but we also used data collected there (tweets) to expose broader activity in the surrounding media ecosystem.

Over time, we noted that a similar kind of rumor kept showing up, over and over again, after each of the man-made crisis events—a conspiracy theory or “alternative narrative” of the event that claimed it either didn’t happen or that it was perpetrated by someone other than the current suspects.

We first encountered this type of rumor while studying the Boston Marathon bombings in 2013. We noticed a large number of tweets (>4000) claiming that the bombings were a “false flag” perpetrated by U.S. Navy Seals. The initial spread of this rumor involved a “cascade” of tweets linking to an article on the InfoWars website. At the time, our researchers did not know what InfoWars was, but the significance of that connection became clear over time.

In subsequent crisis events, similar rumors appeared. After the Umpqua Community College shooting, a rumor claimed the event was staged by “crisis actors” for political reasons—specifically to justify legal restrictions on gun rights. And after the shootings at the Orlando Pulse nightclub, a rumor suggested they were committed by someone other than the accused gunman—with the purpose of falsely blaming the

attack on Muslims. For every man-made crisis event we studied, we found evidence of alternative narratives, often shared by some of the same accounts and connected to some of the same online sites.

These rumors had different “signatures” from other types of rumors. In terms of volume (measured in tweets per minute), most crisis-related rumors spike quickly and then fade out relatively quickly as well, typically “decaying” at an exponential rate. But these alternative narrative rumors rose more slowly, and then they lingered, ebbing and flowing over the course of days or weeks (or years). They also had sustained participation by a set group of Twitter users (i.e. many tweets per user over an extended period of time), rather than finite participation by a large number of users (one or two tweets per user, all at around the same time) as typical rumors do. Additionally, alternative narrative rumors often had high “domain diversity”, in that tweets referencing the rumors linked to a large number of distinct domains (different websites), including alternative media sites such as InfoWars, BeforeItsNews, and RT (aka Russia Today). Several of these rumors also had a strong “botnet” presence—in other words, many participating Twitter accounts were not “real” people, but were operated by a computer program that controlled a large number of accounts.

In our very first study (about the 2013 Boston Marathon Bombings) we noted that alternative narrative rumors intersected with politicized content. Analysis of co-occurring hashtags showed that #falseflag often appeared in the same tweets as #obama, #nra, #teaparty, #tcot, #tlot, #p2. As a researcher of crisis informatics, I've often noted how crises become politicized in online spaces (and elsewhere), but this was different, as the false flag rumor appeared to be deeply connected to political themes and propagated for a distinctly political purpose.

Strange Commonalities and Connections: Why We Shifted Focus

Initially, we chose not to dwell on these types of rumors, thinking that they had little impact on our core research questions—how people respond to crisis events and how we could make the information space more useful for crisis-affected people by detecting false rumors. These alternative narrative rumors rarely resonated within crisis-affected populations. And so, though we often remarked upon them when they surfaced in our data, we maintained our research focus elsewhere.

However, in early 2016, in the wake of the Umpqua Community College shootings and the coordinated terror attacks in Paris, a few of my students decided to take a closer look at what they perceived to be commonalities in the alternative narratives spreading on Twitter about the two different events—as well as what they thought to be a botnet driving a large portion of that content.

[Both of these hunches turned out to be true. The botnet was connected to “the Real Strategy” or TheRealStrategy.com. They coordinated hundreds of accounts that tweeted content related to several different alternative narratives from these events and others. Though some of those accounts have been deleted, others are still operational, new ones have been created, and they continue to publish and tweet out content related to numerous conspiracy theories.]

Using Twitter data collected during these events, the students built network graphs that revealed connections between different Twitter accounts—and between different “communities” of accounts—participating in these alternative narratives. When we went to examine the data in Winter 2016, we were extremely confused by some of the intersections. Why were a handful of “Anonymous” accounts and GamerGaters connected with Pro-Palestinian accounts on one side and European white nationalists on another? Why were seemingly left-wing supporters of Wikileaks connecting with seemingly right-wing supporters of Donald Trump? And why did these groups come together to talk about alternative narratives of mass shooting events? It didn’t make sense. Yet.

A Systematic Exploration of the Alternative Media Ecosystem through the Lens of Alternative Narratives of Mass Shooting Events
Almost a year later, motivated by the political disruptions of 2016, the rhetoric around “fake news” and alternative media, and this nagging feeling that there was something in our online rumorizing data that could provide insight into these issues, we completed a systematic study of alternative narratives of mass shooting events, looking specifically at the alternative media ecosystem that generates them and supports their spread. A first paper resulting from this work was recently reviewed and accepted to the ICWSM 2017 conference. I have uploaded a pre-print version of this paper to my website.

In the remainder of this blog, I am going to describe some of that research, including the methods and the main findings. These findings touch on the nature of alternative media, including the presence of (and connections between) conspiracy theories, political propaganda, and disinformation.

Methods of Data Collection and Analysis

On January 1, 2016, our lab launched a Twitter collection focused specifically on shooting events. We kept this collection going for more than nine months, until October 6, tracking on (English) terms including shooting, shootings, gunman, and gunmen. From this collection, we then identified tweets that referenced alternative narratives—i.e. tweets that also contained terms such as “false flag”, “hoax”, and “crisis actor”.

Next, we created a network map of the Internet domains referenced in these tweets. In other words, we wanted to see what websites people cited as they talked about and constructed these alternative narratives, as well as how those different websites were connected. To do that, we generated a graph where nodes were Internet domains (extracted from URL links in the tweets). In this graph, nodes are sized by the overall number of tweets that linked to that domain and an edge exists between two nodes if the same Twitter account posted one tweet citing one domain and another tweet citing the other. After some trimming (removing domains such as social media sites and URL shorteners that are connected to everything), we ended up with the graph you see in Figure 1. We then used the graph to explore the media ecosystem through which the production of alternative narratives takes place.

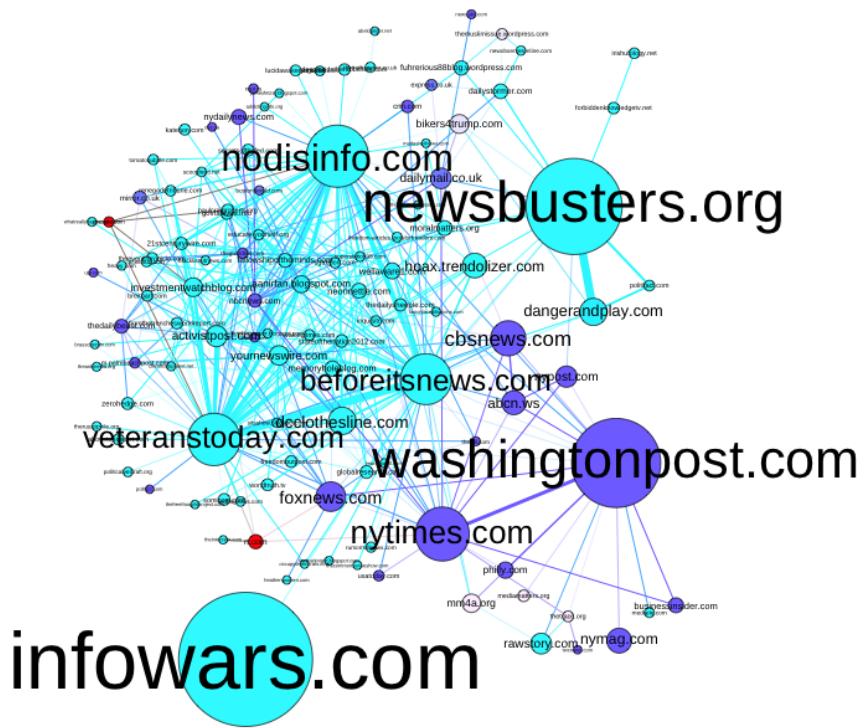


Figure 1. Domain Network Graph, Colored by Media Type
 Purple = mainstream media; Aqua = alternative media;
 Red = government controlled media

After generating the graph, we conducted an in-depth qualitative analysis of all of the domains in the graph—reading their home and About pages, identifying prominent themes in their current website, searching for specific themes within their historical content, examining other available information (online) about their owners and writers, etc. Below, I discuss what we learned about this alternative media ecosystem through this analysis.

Alternative Media Were Cited for Supporting Alternative Narratives; Mainstream Media Were Cited for Challenging Them

The network graph represents a subsection of the larger media ecosystem—it is a snapshot of the “structure” of the conversation around alternative narratives. After trimming to domains cited multiple times (and by multiple people), the graph contains 117 total domains. We determined 80 of these to belong to “alternative media” (Figure 1, colored Aqua) and 27 to belong to mainstream media (Figure 1, colored Purple). Other domains include three belonging to NGOs and

two belonging to media outlets funded by the Russian government (RT.com and SputnikNews.com).

It's important to note that not all of these domains contained content promoting alternative narratives of shooting events. In the Twitter conversations about these alternative narratives, domains were cited in different ways for different kinds of content.

More than half of the domains in the graph (and more than 80% of the alternative media domains) were cited for content explicitly *supporting* the alternative narratives. However, others (especially mainstream media) were cited for factual accounts of the events, and then used as *evidence* by conspiracy theorists as they built these theories. And a few were referenced for their *denials* of these theories. Below are examples of each, to give you a sense of how tweets referenced external domains.

Supporting: The tweet below links to an article in the WorldTruth.tv domain which claims that witness accounts of multiple gunmen (which conflict with the official account) suggest that the Orlando Pulse nightclub shooting is some sort of false flag. Contradictory and dynamic information—typical of the fog-of-war type situations that occur after crisis events—is often used as “evidence” to support alternative narratives of these events.



As Evidence: The tweet below claims that one of the witnesses to the Orlando shooting is an actor and that the shootings were a false flag. This echoes a common theme, which appears across many alternative narratives in our research, that “crisis actors” are used to stage events. The tweet links to an article in the Toronto Star domain which contains a neutral, factual account of the event.

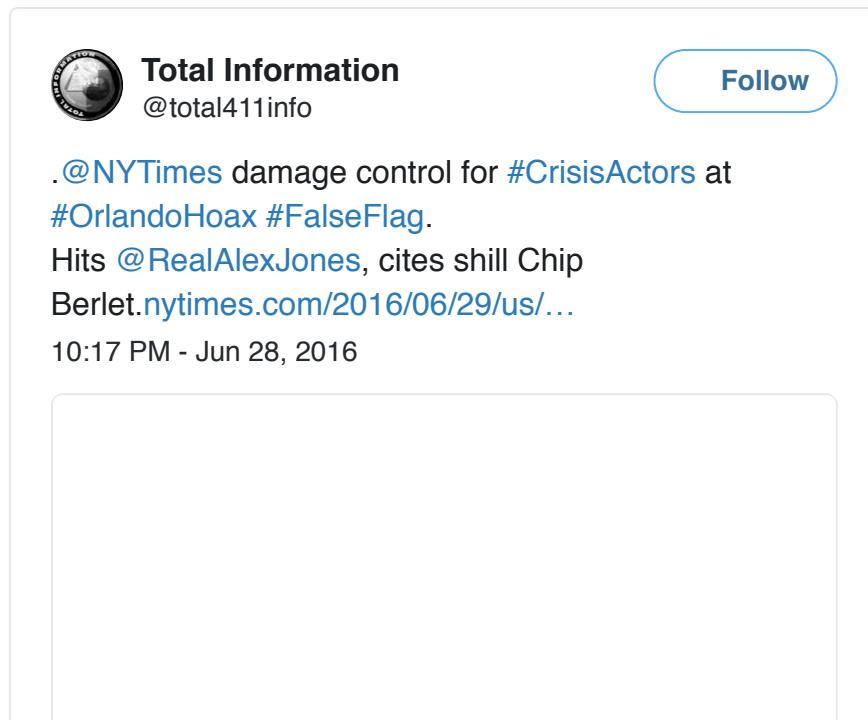
Adolf Otto Eichmann
@Kage43Kent

Follow

#Actor Luis Burbano #Orlando Shooter called 911 three times before the killing thestar.com/news/world/201...
#Falseflag twitter.com/jcat111/status...

11:52 AM - Jun 15, 2016

Denying: This tweet links to the New York Times domain—to an article that refutes several different alternative narratives of the Orlando shootings. However, instead of aligning with the arguments in that article, this tweet is accusing the New York Times of being a participant in the conspiracy/hoax/false flag.



[Following Twitter's rules, I am only providing examples here of tweets that are still publicly available on Twitter. I have also attempted to choose accounts for these examples that seem to intentionally propagate alternative narratives—in other words, I am attempting to avoid calling out individuals/accounts that might be uncomfortable being associated with these ideas.]

Most of the domains cited in the production of alternative narratives were “alternative media” domains, and most of these (68 of 80) were cited (linked-to) in the tweets we collected for content that explicitly supported alternative narratives. As you can see in the graph (Figure 1), the alternative media ecosystem is tightly connected—i.e. the Twitter users who produce alternative narratives often cite several different alternative media domains in their conspiracy theory tweets. The three main hubs in this particular network are VeteransToday.com, BeforeItsNews.com, and NoDisinfo.com, but there are many other alternative media domains that play a significant role in the production of alternative narratives. This alternative media ecosystem (a subset of the larger graph) is the focus of the remainder of this blog.

However, I want to explicitly note and clarify one aspect of the graph: though mainstream media domains like the Washington Post, the New York Times, and Fox News appear in the graph, **no mainstream media account in this graph hosted any content promoting the alternative**

narratives we were studying. Instead, they were typically cited in our Twitter data for general content about the event that was later used as “evidence” of a conspiracy. Mainstream media were also cited for corrections of the alternative narratives (sometimes in tweets supporting those corrections, sometimes in tweets contesting them). In the case of the New York Times, the newspaper posted an article explicitly denying alternative narratives of the Orlando shooting event. This denial was then cited several times by those promoting those narratives—as even more evidence for their theory. [*This demonstrates a vexing aspect of rumor-correcting in this context—that corrections often backfire.*]

The network graph does reveal some mainstream media sites to be more integrated into the alternative media ecosystem. For example, several people who tweet links to VeteransToday.com also tweet links to FoxNews.com, pulling it closer into that part of the graph.

The Role of Botnets in Amplifying Alternative Narratives

These data also provide insight into the effect of automated accounts (botnets) on the data. For example, the most tweeted domain in our data was TheRealStrategy.com. It was tweeted so many times (7436) and connected to so many domains (relative to all other domains) that we had to remove it from the graph. [*It was the only highly cited, highly connected media domain we removed.*] Examining the temporal patterns (tweets over time) suggests that almost all of the tweets that linked-to this domain were generated by a computer program. That program operated hundreds of different accounts, directing them to tweet out in regular bursts (dozens at the same time). Most often, these tweets linked to TheRealStrategy, but the program also sprinkled in tweets linking to other alternative media domains. Closer analysis revealed many of these Twitter accounts to have similar profile descriptions and to use photos stolen from other people online. This is a very sophisticated botnet that seems to be effectively bringing “real” accounts into its friend/following networks—and primarily propagating conspiracy theories and politicized content.

The InfoWars site was the second-most highly tweeted in our data set (1742 times). Almost all of the tweet activity citing InfoWars came from a coordinated set of accounts—all were similarly named and each

sent a single tweet linking to one of two InfoWars articles about different alternative narratives of different shooting events. All of these accounts are now suspended. Though not as sophisticated as TheRealStrategy, this botnet did amplify the content of InfoWars, which was occasionally picked up and retweeted by others.

Political Propaganda: Nationalism vs. Globalism

One of the first things that struck us as we conducted qualitative content analysis on the alternative media domains was the amount of political content on the websites. We attempted to characterize this content, going through several rounds of iteration to try to recognize patterns across the sites and distinguish between different political orientations.

It quickly became clear that the U.S. left (liberal) vs. right (conservative) political spectrum was not appropriate for much of this content. Instead, the major political orientation was towards anti-globalism. Almost always, this orientation was made explicit in the content.

The meaning of globalism varied across the sites. For some websites focused on a U.S. audience, globalism implied a pro-immigrant stance. For more internationally-focused sites, globalism was used to characterize (and criticize) the influence of the U.S. government in other parts of the world. In some of the more conspiracy-focused sites, the term was used to suggest connections to a global conspiracy by rich, powerful people who manipulated the world for their benefit. Globalism was also tied to corporatism—in other words, the ways in which large, multi-national companies exert power over the world. And the term was also connected, implicitly and explicitly, to mainstream media.

In this way, to be anti-globalist could include being anti-mainstream media, anti-immigration, anti-corporation, anti-U.S. government, and anti-European Union. Due to the range of different meanings employed, the sentiment of anti-globalism pulled together individuals (and ideologies) from both the right and the left of the U.S. political spectrum. Disturbingly, much of the anti-globalist content in these alternative media domains was also anti-Semitic—echoing long-lived

conspiracy theories about powerful Jewish people controlling world events.

So Many Conspiracy Theories: Crippled Epistemologies, Muddled Thinking, and the Fingerprints of a Disinformation Campaign

Another thing we noticed was both a proliferation and a convergence of different conspiratorial themes. Every domain that hosted an article promoting an alternative narrative of a shooting event also contained content referencing other conspiracy theories—sometimes hundreds of them. They were not all political in nature. We also encountered pseudo-science theories about vaccines, GMOs, and “chemtrails”. Some domains were all about conspiracy theories, but others featured seemingly normal news with conspiracy theories sprinkled in. Through qualitative analysis, we determined 24 alternative media domains to be primarily focused on distributing conspiracy theories and 44 to be primarily focused on communicating a political agenda.

Though there were many different theories spreading through this information ecosystem, we also saw a convergence of themes—some of the same stories appeared on several different domains. Occasionally, the stories seemed largely independent (i.e. different perspectives, different evidence), but often they were essentially copied from one site to another, or a downstream story simply synthesized an article on another site, including lengthy excerpts from the original. Additionally, a few authors seemed to contribute stories to multiple domains in the network.

So, a person seeking information within this ecosystem might encounter an article from one website that synthesized an article from a second website that was originally posted on and copied from a third website. One effect of this is that people seeking information within this space may think they are getting information from many different sources when in fact they are getting information from the same or very similar sources, laundered through many different websites. Sunstein & Vermeule (2009) argue that conspiratorial thinking is related to a “crippled epistemology” and that a significant component of this is a limited and/or slanted information diet (for example, one shaped by a

social group). Our research suggests the information dynamics of this alternative media ecosystem, how the same information exists in different forms in different places, may create a false perception of information diversity or triangulation—further complicating this issue of crippled epistemologies.

From another perspective, these properties of the alternative news ecosystem—the proliferation of many and even conflicting conspiracy theories and the deceptive appearance of source diversity—may reflect the intentional use of disinformation tactics. Though we often think of disinformation as being employed to convince us of a specific ideology, in a 2014 article titled “The Menace of Unreality”, Pomerantsev and Weiss describe how Russian disinformation strategies (which they trace back to Lenin) are designed not to convince but to confuse, to create “muddled thinking” within society. Their strategic argument is that a society who learns it cannot trust information can be easily controlled. It is possible that the current media ecosystem—including the alternative media domains and the social media platforms that help spread links to these domains—is contributing to muddled thinking (a relative or effect perhaps of an crippled epistemology). It is not yet clear if these effects are related to purposeful disinformation campaigns or are just emergent effects of our current information space. It seems researchers have some work to do to both clarify what is happening here and to perhaps think about designing systems that are more resilient to disinformation.

Alternative Media Co-opt Critical Thinking, Facts, and Truth

Perhaps the most vexing finding that emerged from this analysis—especially as we attempt to think of how to help people become better consumers of online information—was what we perceived to be an intentional strategy by many alternative media websites to leverage rhetoric around fake news and critical thinking to further confuse and mislead readers.

Our research shows that rejection of mainstream news is a common theme across alternative media domains. Perhaps it's a truism to say that alternative media exist in juxtaposition to mainstream media, but what is interesting here is that many alternative media sites have explicitly set themselves up as opposition to mainstream, “corporate”

media. They have also seized upon claims of political bias in mainstream media (towards liberal or pro-Western ideologies) and have leveraged those to support their own legitimacy.

Additionally, it seems they have co-opted arguments about media literacy ([boyd makes this same argument](#)) and critical thinking. The conversation around “fake news” often ends with statements about teaching people to become better consumers of information—to be skeptical as they educate themselves through encounters with online media. Alternative news sites have appropriated these arguments and are using them to support the propagation of alternative narratives and other conspiracy theories.

Consider the text below, an excerpt from the About page of the 21stCenturyWire.com domain:

The 21st Century is the beginning of a new information epoch and you, the reader, are the freshman class of free and critical thinkers in a new and dynamic information age. In its totality, there is an immense volume of information to be had on the Internet but we truly believe that in this new decentralised, grassroots and egalitarian World Wide Web, the cream will eventually rise to the top – but this can only be achieved by keeping this internet ‘neutral’ and free from excessive government and corporate control.

21stCenturyWire.com is a typical domain in our network graph, positioned in the upper left corner (of Figure 1) and strongly connected to both NoDisinfo and VeteransToday (which both spread strong anti-Semitic content). 59 tweets in our collection linked to this domain, referencing multiple articles explicitly supporting alternative narratives about several mass shootings, including claims that both the Dallas police shootings and the Orlando nightclub shootings were staged events. However, the conspiratorial focus of this domain extended far beyond alternative narratives of shootings. Domain content supported a wide range of conspiratorial themes, with articles promoting claims about vaccines causing autism, government-engineered weather events, George Soros-backed anti-Trump protests, and pedophile rings operated by powerful people. Through our analysis of domain content, we also determined 21stCenturyWire to be strongly supportive of Russian political interests (another prominent theme in our data).

The domain is owned and operated by Patrick Henningsen, a journalist who has worked for RT news, Guardian.co.uk, GlobalResearch.ca, and

Infowars.com. Perhaps not surprisingly, all of these domains are nodes in our graph.

Examining the About page of 21stCenturyWire, you can see how the site leverages the (somewhat techno-utopian) rhetoric of freedom of information and citizen-journalism—explicitly encouraging readers to use their own “critical thinking” skills while implicitly complimenting them on those skills and perhaps activating a sense of confidence in their abilities. *You can handle this. We'll give you the facts and you can decide for yourself!* The site also claims to be outside both corporate and government control. The first claim represents a somewhat natural counter-positioning—i.e. alternative media against corporate-controlled mainstream media. But the second claim is somewhat disingenuous, as the domain often hosts content that is cross-posted to RT—formerly Russia Today, a media outlet funded and largely controlled by the Russian government.

This kind of positioning of alternative media was typical for the domains we examined. Below is another example, this one from the Purpose & Goals page of the NoDisinfo.com domain:

Our mission is to provide information that isn't twisted. To provide pure facts. To provide intelligent commentary on the facts. To never distort the facts. This is so people can make up their own minds, without corruption. Another purpose is to unravel deception and disinformation, so that a true understanding can be achieved. This is so people can compare the false information with real and uncompromised information.

It's just pure truth. If we make a mistake in real truth, let us know. We'll correct it. That's why it's called nodisinfo.

Thus, this site is committed only to facts. This may include theories based on these facts. Feel free to add to these truths and factual data.

Comments are accepted. But the use of foul language and slander is not allowed.

Notice the language emphasizing how this website provides “facts”. It allows people to “make up their own minds”. Its purpose is to unravel “deception and disinformation”. This framing is likely very intentional, claiming to be presenting unadulterated “truth” and empowering users to perhaps feel that they are discovering that truth within this domain. And users can find all kinds of truth (in the form of conspiracy theories) here—from 9–11 trutherism to claims about possibly apocalyptic effects of the Fukushima nuclear disaster being purposefully obscured by mainstream media.

Summary and Conclusion

This research attempted to take a systematic approach to unpacking the alternative media ecosystem. We focused on “alternative narratives” of crisis events and utilized Twitter data to map the structure of the alternative media ecosystem that drives these narratives. Through content analysis, we found these domains to collectively host many different types of conspiracy theories—from politically-themed narratives about the “New World Order” to anti-vaccine arguments. In this “virtual” world, the Sandy Hook School shootings were staged by crisis actors and the earth is actually flat after all.

We determined a large portion of the content on this network to be political propaganda. For the most part, this political propaganda was focused around “anti-globalism”. This term was used to designate different things in different domains (and even in different articles within the same domains)—e.g. anti-immigration, anti-Western imperialism, anti-corporation, anti-media. Disturbingly, there were also strong currents of antisemitism (sometimes explicit, sometimes less so) across a subsection of this ecosystem. Taken together, these positions seem aligned with and used in support of the rise of nationalist ideologies in the U.S. and elsewhere.

We also noted how the structure of the alternative media ecosystem and the content that is hosted and spread there suggest the use of intentional disinformation tactics—meant to create “muddled thinking” and a general mistrust in information.

Because the underlying data in this analysis are limited (to tweets about shooting events), future work will be needed to A) assess the broader alternative media ecosystem (our data limited us to a very specific view); and B) determine how influential these media and their messages are on U.S. and global perspectives of world events and science. However, it is clear that information shared within this seemingly fringe information ecosystem is entering the public sphere at large.

When we conducted this analysis in December, many of these alternative news domains were beginning to appropriate the term “fake news” to deflect attacks back onto the mainstream media. Weeks later, newly inaugurated U.S. President Trump echoed this refrain, publicly stating (even tweeting) that various mainstream media outlets and particular stories were “fake news”. Other information trajectories from

alternative media websites to public statements by the Trump administration have been identified (e.g. the recent wiretapping claims), and though this does not imply causation, it does indicate a connection between the alternative media ecosystem and the U.S. President. The addition of Steve Bannon to Trump's inner circle underscores this connection as well. Before his appointment to Trump's campaign, Bannon ran Breitbart news, an alternative media website that appears in our data—and one that we determined to have a strong anti-globalist perspective. Indeed, Bannon's recent comments at the Republican CPAC meeting make this ideological orientation explicit.

While criticizing the mainstream media, Bannon said this: "They're corporatist, globalist media that are adamantly opposed to an economic nationalist agenda like Donald Trump has."

This comment summarizes a great deal of the research we did, demonstrating how criticism of mainstream media (practically etched into the DNA of alternative media) is aligned with a political agenda of anti-globalism in favor of nationalism, and how that agenda is connected to the political orientations and goals of the Trump administration. Perhaps the main contribution of our research is merely to point out that these ideologies are spread within an alternative media ecosystem that utilizes conspiracy theories like Sandy Hook hoax claims and old anti-Semitic narratives to attract readers and support this spread. And that these alternative media websites aren't focused solely on U.S. far-right or alt-right content, but are also using alt-left content to pull readers into this information ecosystem and the ideologies spreading there.

Most importantly, this work suggests that Alex Jones is indeed a prophet. Seriously, as I read through dozens of these alternative media websites and dug DEEP into their content, I realized that there is an indeed an information war being waged. Three years ago, our lab decided these conspiracy theories were too marginal and salacious to be the focus of our research. Almost that it was beneath our dignity to pay attention to and promote this kind of content. What a terrible mistake that was. It seems to me that we weren't the only ones who made it. It is (past) time we attend to this (as researchers and designers of the systems that conduct this content). I hope it is not too late.

[Here is a list of the domains that appear in our network graph. Please note that the qualitative coding was done through iterative, interpretive content analysis. It is possible that others may perceive that a different determination (or set of categories) would be better for some of these domains. Please let me know if you feel that there is a systematic coding error or unrecognized pattern in the data, as this work is ongoing and I'd love to be able to incorporate your insights. Thank you.]



Prof. Dr.
Bernd Holznagel, LL.M.

THE PAPER WILL BE PRESENTED AT THE CENTRE D'ÉTUDES SUR LES MÉDIAS (CEM),
MONTREAL - 05.10.2017

NETWORK ENFORCEMENT ACT – THE GERMAN APPROACH TO FIGHT FAKE NEWS, DIGITAL VIOLENCE AND TERRORIST SPEECH IN SOCIAL NETWORKS

I. Dimensions and implications of hate communication

The growing dissemination of hate crime and other criminal contents, especially in social networks such as Facebook, YouTube and Twitter, gave rise to the Act to Improve Enforcement of the Law in Social Networks of September 7th, 2017 (NEA)¹. According to the figures published by the Federal Office of Justice in 2016, the number of hate crimes on the Internet in 2015 has nearly tripled compared to the previous two years.² Surveys also show that a lot of things remain unknown. In a representative online survey³ of German citizens from June 2016, two-thirds of the users stated that they had already been confronted with hate messages in

The author thanks Maximilian Hemmert-Halswick for preparing the manuscript and for translation work.

¹ BGBI. I S. 3352. For the explanatory statements see BT-Drs. 18/13013, BR-Drs. 536/17.

² Siehe https://www.bundesjustizamt.de/DE/Themen/Buergerdienste/Justizstatistik/Straftaten/Strafrechtspflege_node.html

³ Forsa, Ehtik im Netz, Hate Speech, 2016, http://www.lfm-nrw.de/fileadmin/user_upload/lfm-nrw/Service/Veranstaltungen_und_Preise/Med.ienversammlung/2016/EthikimNetz_Hate_Speech-PP.pdf.

social networks, Internet forums, or blogs. In that survey, “hate speech” was defined as comments aimed against a specific person or a specific group of people due to their ethnic or religious affiliation, their national origin, sex, age, disability or sickness, that included statements of hatred, threats of or incitements to violence. In the American discussion, this is referred to as hate speech, digital violence, or terrorist speech.⁴ Looking at a group of 14- to 24-year-old adolescents, 91 percent mentioned corresponding experiences. Roughly every third respondent felt intimidated by such comments.

The effects do not only account for the “average” user of social media. A survey published in April 2017 by the Council of Europe, based on a sample of 940 journalists reporting from the 47 member states of the Council of Europe and Belarus, revealed that professionals in the field of journalism are also affected. Fear of psychological violence (60%), cyberbullying (57%) and intimidation by individuals (51%) or interest groups (45%), and even physical violence (41%) influences journalistic work and leads to self-censorship. Many journalists felt compelled to tone down controversial stories (31%), withhold information (23%), or abandon stories altogether (15%).⁵ Furthermore, online harassment has a clear gender-specific orientation and is particularly aimed at minorities. In the year 2016, the daily newspaper, “The Guardian”, commissioned an investigation of the comments made on its website since 2006. Regardless of the concrete content, articles written by female journalists drew more hate comments than those written by male counterparts. Among the ten most bothered authors were eight women and two black men.⁶ Obviously, this can have a deterrent effect on Internet communication and journalism.

⁴ Tsesis, “Terrorist Speech on Social Media, 70 Vand. L. Rev. 651 (2017).

⁵ Clark/Grech, *Journalists under Pressure. Unwarranted interference, fear and self-censorship in Europe*, 2017,
<https://wcd.coe.int/ViewDoc.jsp?p=&id=2456911&Site=DC&BackColorInternet=F5CA75&BackColorIntranet=F5CA75&BackColorLogged.=A9BACE&direct=true>.

⁶ Gardiner et al., The dark side of Guardian comments,
<https://www.theguardian.com/technology/2016/apr/12/the-dark-side-of-guardian-comments>.

The Ministry of Justice concludes that "if hate speech and terrorist speech are not combated properly, they pose a massive threat to peacefully living in a free, open and democratic society".⁷

Apart from that, the experiences of the United States 2016 presidential election campaign⁸ have lead Germany to also prioritize the combat against punishable "Fake News" in social networks.⁹

II. Background of the NEA

In the run-up to the legislative process, various ways of dealing with these developments were considered on the European stage. In light of the terrorist attacks in Paris (2015) and Brussels (2016), which apparently were also organized through social networks,¹⁰ the European Commission felt compelled to make concrete proposals. Last year, she and large IT companies announced the "Code of Conduct on Countering Illegal Hate Speech Online"¹¹. This is not a law (regulation or directive), but a non-binding "soft law". Most prominently, this code stipulates that the deletions of illegal content should be performed by the network operators within 24 hours of receipt of the complaint.

The EU was only able to bring the big American social media platforms to this agreement by announcing that otherwise binding and enforceable laws would be adopted.¹² Said Code of Conduct appeared to be the lesser of two evils.

Shortly after this, in the fall of 2016, in Germany, a task force composed of representatives of the major social networks and members of the government,

⁷ NEA draft of the Federal Ministry of Justice, March 27, 2017, p. 10, accessible in English: <http://ec.europa.eu/growth/tools-databases/tris/de/index.cfm/search/?trisaction=search.detail&year=2017&num=127&mLang=de&CFID=2665602&CFTOKEN=e657eec98ea2b052-AE96FBFC-B2FB-F82B-D08C114CC7B379C1>, S. 10.

⁸ For example: https://www.buzzfeed.com/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook?utm_term=.sxl44PBgJn#.va7GGEveQ0 or <http://www.cbsnews.com/news/fake-news-macedonia-teen-shows-how-its-done/>.

⁹ BT-Drs. 18/13013, p. 1.

¹⁰ *Toor*, UK Lawmakers Say Facebook, Google, and Twitter are 'Consciously Failing' to Fight ISIS Online, THE VERGE, August 26, 2016), [http://www.theverge.com/2016/8/26/12656328/facebook-google-twitter-isis-propaganda-uk-report; *Torok*, How social media was key to Islamic State's attacks on Paris, <http://theconversation.com/how-social-media-was-key-to-islamic-states-attacks-on-paris-50743>.](http://www.theverge.com/2016/8/26/12656328/facebook-google-twitter-isis-propaganda-uk-report)

¹¹ http://ec.europa.eu/justice/fundamental-rights/files/hate_speech_code_of_conduct_en.pdf.

¹² *Keats Citron*, „Extremist Speech and Compelled Conformity”, University of Maryland Francis King Carey School of Law Legal Studies Research Paper No. 2017-12, p. 16.

appointed by Federal Justice Minister Heiko Maas, formulated similar objectives. Namely, a more transparent handling of complaints, faster deletion of unlawful content, and an overall better cooperation between networks and authorities. However, the networks did not meet the requirements of the Code of Conduct nor the objectives of the national task force: only 39 percent of the complaints at Facebook led to the removal of the reported content, as for Twitter the removal rate is 1 percent. Both social networks needed more than 24 hours to deal with complaints in half of all cases.¹³

In December 2016, large media attention came to Facebook's handling of a complaint by the chairwoman of the German Bundestag's Legal Committee, Renate Künast (Green Party). On Facebook, a picture with a reputed quotation by Ms. Künast was spread widely, which alluded to the alleged murderer. According to this, Ms. Künast said: "The traumatized young refugee has killed, but we must help him now." The well-known newspaper "Süddeutsche Zeitung" was headed as the source of that statement. The quotation and source were both fictitious. This is a typical case of "Fake news", as it was often used in the American election campaign against Hillary Clinton. Although the post was relatively easy to identify as such, Facebook did not react to the complaint of the popular former Consumer Protection Minister. Only after numerous high-ranking MPs had expressed their misunderstanding about this practice, Facebook came to terms with their deleting obligation.

This process, which received wide attention in the German public, accelerated the project to set up a law to improve the law enforcement on the Internet. The NEA has since then been called "Facebook Law".

On June 30, 2017, the German Bundestag voted to ratify the NEA with the votes of ruling Grand Coalition. It will enter into force on October 1, 2017. The European Commission, which had been notified in application of Directive 2000/31/EC ("E-Commerce-Directive")¹⁴, raised no objections to the law.¹⁵ The Federal Council has

¹³ BR-Drs. 315/17 p. 2.

¹⁴ Notified. Draft Law as of March 27, 2017, access via: <http://ec.europa.eu/growth/tools-databases/tris/de/index.cfm/search/?trisaction=search.detail&year=2017&num=127&mLang=DE>.

¹⁵ Though some argue that the NEA violates European law: *Spindler*, ZUM 2017, 473; *Hoeren*, <https://community.beck.de/2017/03/30/netzwerkdurchsetzungsgesetz-europarechtswidrig>.

proposed to assign the State Media Authorities with the enforcement of the NEA, but it did not, in principle, question the Federal Government's competence to make regulations in this field.¹⁶

III. The approach of the NEA

The aim of the NEA is to ensure that social networks increasingly fulfill their duty to delete illegal content. The NEA does not create any new deletion obligations. Rather, it presupposes them. This is often misunderstood in the public debate.

The obligation not to disseminate unlawful contents of third parties arises directly from the various criminal provisions and from Sec 1004 of the German Civil Code, which stipulates the elimination and omission of an impairment of certain rights (e.g. the right of personality and portrait rights). Ms. Künast, for example, could state an infringement of both. There is one important exception to this rule. Host providers are only liable for the support they provide to third-party crimes (as in uploading illegal content) if such content has been notified to them. Then they must delete it immediately. This so-called Notice-and-Take-Down principle has been laid down in the E-Commerce Directive 2000 (Article 14¹⁷) and has been transposed into German law by Sec 10 Tele Media Act.

This “provider privilege” is not God-given, but a decision of the legislator, who wants to promote certain business models on the internet. The NEA wants to form the legal framework for this. In doing so, the lawmakers must consider the impact of fundamental rights. According to the jurisprudence of the Federal Constitutional Court, fundamental rights are valid for all areas of the law due to their objective function (“radiating effect”). This also applies to the legal relations between private

¹⁶ Vgl. BR-Drs. 315/17 p. 8. Some argue that the federal legislator has no competence: *Hain/Ferreau/Brings-Wiesen*, K&R 2017, 433 (434 f.); *Gersdorf*, MMR 2017, 439 (440 ff.).

¹⁷ Art. 14 E-Commerce-Directive: “Where an information society service is provided, that consists of the storage of information provided by a recipient of the service, Member States shall ensure that the service provider is not liable for the information stored, at the request of a recipient of the service, on condition that: (a) the provider does not have actual knowledge of illegal activity or information and, as regards claims for damages, is not aware of facts or circumstances from which the illegal activity or information is apparent; or (b) the provider, upon obtaining such knowledge or awareness, acts expeditiously to remove or to disable access to the information.”

parties, the so-called third-party effect of fundamental rights¹⁸. This principle is of central importance in this context.

As a mere technical service provider, which stores or transmits third-party information to a user in a communication network, the network operators cannot claim freedom of expression for themselves.¹⁹ On the contrary, they provide an infrastructure for free, public communication, which is of paramount importance in a democracy. Thus, they assume functions formerly assigned to the state (general public provisions in the field of telecommunication). For this constellation, the Federal Constitutional Court has emphasized the third-party effect of the fundamental communication rights.²⁰

The protection of freedom of expression on the Internet has two sides. For one side, it is just exercise of freedom of expression. For the other side, hate speech and threats in social networks hinder democratic participation. Minorities are silenced; journalists censor themselves; women are marginalized.

The legislator also has protection obligations against this group, which he must bring into balance with the freedom of expression.²¹ When choosing the regulatory model, he has a margin of appreciation.

The NEA is based on a controlling approach that is successfully being used to enforce legal standards in economic enterprises and in the financial market.²² There, companies themselves are required to monitor compliance with certain standards, and they must implement and report on effective procedures to ensure compliance with the law. In the case of infringements, sanctions can be imposed.²³ The idea behind the NEA is to introduce compliance regulations for social networks.

¹⁸ BVerfGE 7, 198 (205 ff.); 86, 122 (128); 107, 275 (280); 112, 332 (358).

¹⁹ The explanatory statement only mentions an interference with the professional freedom (BT-Drs. 18/12356, p. 22).

²⁰ BVerfG, 1 BvR 699/06 -, juris, pt. 56, 59.

²¹ BVerfGE 89, 214 (232); BVerfG NJW 1998, 1475 (1476); NJW 2011, 3428 (3432); NJW 2014, 46 (47); 2016, 2247 (2248).

²² Bottmann, in: Park (Ed.), Kapitalmarktstrafrecht, 4th ed.. 2017, Chap. 2.1 pt. 1.

²³ Schäfer, in: Park (Ed.), Kapitalmarktstrafrecht, 4th ed.. 2017, Chap. 2.2 pt. 9.

IV. The NEA

1. Social Networks as the regulated bodies

Pursuant to Sec 1 (1) NEA, the provisions of the NEA apply to tele-media service providers²⁴ which, for profit making purposes, operate social networks. Social networks are defined as “internet platforms that enable users to exchange and share any content with other users or to make such content available to the public”.

The provision for the exchange of *any* content leads to the exclusion of networks that are designed to the exchange of *specific* content, e.g. professional networks, online games or sales platforms.²⁵ The term “platform” refers to a communication space that connects many participants.²⁶ Individual communication services, such as e-mail or messenger services, are excluded from the scope of application.²⁷

Platforms with journalistic and editorial contents do not count as social networks, as far as the NEA is concerned. It remains unclear whether journalistic and editorial content, which is offered on social networks, falls within the scope of application. The better arguments, in particular the importance of largely unimpeded reporting, speak against the inclusion.²⁸

In Sec 1 (2) NEA, the applicability is limited to social networks with more than 2 million registered users in Germany. These bigger networks have a significantly higher potential of perpetuating legal violations than smaller networks²⁹, and they can adequately cope with the extra effort associated with the duties of the NEA.³⁰

2. Removing and blocking of unlawful content

The rules for removing of or blocking access to unlawful content are the core elements of the NEA.

²⁴ Tele-med.ia service provider shall mean any natural or legal person who provides his own or third-party tele-med.ia for use or provides access. Tele-med.ia is defined as electronic information and communication services if their service does not only entirely consist in the transmission of signals over telecommunications networks.

²⁵ BT-Drs. 18/13013, p. 20.

²⁶ BT-Drs. 18/13013, p. 21.

²⁷ BT-Drs. 18/13013, p. 20.

²⁸ Statement of the DJV (German Federation of Journalists), p. 4.

²⁹ BT-Drs. 18/13013, p. 21.

³⁰ Furthermore, the reporting obligations only accounts for networks that receive more than 100 complaints about illegal content in one calendar year.

a) Removal within 24 hours

A “manifestly” unlawful content must be removed by the network operator within 24 hours after receiving the complaint. Pursuant to Sec 1 (3) NEA, “unlawful content” is defined as content on social networks, which fulfils the requirements of the [22 enlisted offences of the German Criminal Code]”. The enlisted offenses³¹ can be divided into provisions on State security (e.g. Sec 129a GCC “forming terrorist organizations”; Sec 91 GCC “encouraging the commission of a serious violent offence”), offenses against public order (these offences generally require the suitability for disturbing the public peace, e.g. Sec 130 GCC “incitement of hatred”) and provisions on the protection of the personal honor (e.g. Sec 185 GCC “insult” or Sec 187 GCC “intentional defamation”). In the draft law, three more offences were included that have not been adopted in the final version; these were offences against the reputation of the state (e.g. Sec 90 “defamation of the President of the Federation”). A content is manifestly unlawful if it can be identified as such by trained personnel within 24 hours. If there remains any doubt in fact or in law, the content is not manifestly unlawful.³²

The common ground of these offences is that they serve to guarantee a non-violent communication space for all citizens. In some cases “fake news” must be removed. For example, it is unlawful to asserts or disseminates intentionally and knowingly an untrue fact related to another person, which may defame him or negatively affect public opinion about him (Sec 187 GCC). Under a human rights perspective this is not an issue because untrue facts do not fall into the scope of freedom of expression.

b) Removal within 7 days

Every other unlawful content must be removed within seven days after receiving the complaint. This extended time frame enables the network to obtain a statement from the author or external expertise.

³¹ Sec 86, 86a, 89a, 91, 100a, 111, 126, 129 bis 129b, 130, 131, 140, 166, 184b in conjunction with 184d, 185 bis 187, 201a, 241 or 269 of the German Criminal Code.

³² BT-Drs. 18/13013, p. 22.

The time limit of seven days may be exceeded in certain, more complex cases,³³ e.g. if the decision on unlawfulness depends on the truth or falsehood of individual allegations. The removal can also happen later than seven days after the complaint if the social network transfers the decision on the illegality to a Regulated Self-Regulation³⁴ facility.

3. The Procedure for handling complaints and reporting obligations

Pursuant to Sec 3 (1) NEA, the procedure must ensure that the provider of the social network immediately takes note of the complaint and checks whether the content is unlawful and – if unlawful – removes or blocks access to it.

The NEA requires the complaint itself to be sufficient to examine the unlawfulness.³⁵ A complaint which does not allow a qualified examination is, in principle, irrelevant.³⁶ In this case, however, whether the complainant must be given the opportunity to further substantiate his/her complaint remains unanswered. The networks will have to find solutions to this question, which is important in practical terms.

In addition, the network operator must promptly inform both the author and the person submitting the complaint of any action taken and must also provide them with reasons for its decision. This regulation is intended to allow the persons affected to take a targeted approach against the decision.³⁷

Sec 2 NEA sets forth the reporting obligations of network operators. They must report to the competent authorities every six months³⁸ in German on how they implement the guidelines of the NEA. Primarily, operators are to provide information on the number of complaints and deletions made. This illustrates how important it is for the legislator to show results.

³³ BT-Drs. 18/13013, p. 22.

³⁴ See IV.4.

³⁵ BT-Drs. 18/13013, p. 22.

³⁶ BT-Drs. 18/13013, p. 22.

³⁷ BT-Drs. 18/12356, p. 23.

³⁸ The draft law stipulated the report to be due every three months (BT-Drs. 18/12356).

4. Regulated Self-Regulation

A network operator may transfer the assessment of contents to a facility of Regulated Self-Regulation.³⁹ Regulated Self-Regulation is referred to when monitoring or compliance purposes non-governmental institutions are integrated in a governmental framework.⁴⁰ Self-regulatory institutions are usually organized by the regulated companies themselves. The state controls the self-regulatory bodies through the means of recognition and, if necessary, the revocation of this recognition. Regulated Self-Regulation is – at least in Germany – mainly known from youth media protection, which served as a role model for the NEA in this aspect.⁴¹ The most commonly known institution of Regulated Self-Regulation is probably the voluntary self-control of the film industry (FSK). It is responsible for compliance with youth protection law in the film sector.

For an institution to be recognized as a facility of the Regulated Self-Regulation, Sec 3 (6) NEA states that the inspectors at the institution must be independent and have the required expertise. An expeditious inspection of a content within seven days must also be ensured. It must provide rules of procedure that regulate the scope and course of the inspection, which subsequently enables decisions to be reviewed. A complaints center must be set up. Lastly, the NEA requires that the facility is organized by several social networks.⁴² It must be noted that co-organizing or joining the Regulated Self-Regulation is a completely voluntary decision by each network.

5. Person authorized to receive service in Germany

According to Sec 5 NEA, the operators of social networks are obliged to appoint a person authorized to receive service, and indicate that person on their platforms. The purpose of this provision is to enable notifications in fines proceedings pursuant to Sec 4 NetzDG or court proceedings before German courts for the dissemination of unlawful content. If the prosecution seeks information from him, he must respond within 48 hours. This rule also applies to foreign companies, as the

³⁹ Sec 3 (2) No. 3b NEA.

⁴⁰ Schulz/Held, in: Hahn/Vesting (Ed.s.), Beck'scher Kommentar zum Rundfunkrecht, 3 rd. ed.. 2012, § 1 JMStV, pt. 22.

⁴¹ BT-Drs. 18/13013, p. 23.

⁴² Sec 3 (6) No. 1 - 5 NEA.

enforcement of the law can be very difficult in cases with a foreign dimension; the networks simply did not answer certain requests.

6. Fines

The NEA provides for penal sanctions of up to EUR 50 million for the intentional or negligent breach of the obligations of the NEA. Fines are not imposed for singular violations of the deleting obligation.⁴³ Fines can only be imposed if there are “systemic flaws”, which means that the network operator has failed to implement a functioning procedure for handling complaints. In short, the breach must exhibit an exceptional seriousness. There are no further criteria of what accounts for a “systemic flaw”. Hence, the implementation of this regulation will be very difficult in practice.

V. Evaluation

1. Approval of the objectives and the regulatory approach

The goal of the NEA has been welcomed in public.⁴⁴ The chosen regulatory model has also received broad approval. A compliance-based approach has proved successful to regulate complex organizations in civil society. The possibility of transferring deleting decisions to a facility of self-regulation is thought to be a good idea. It has been included in the draft law only after the hearings of the Legal Committee and it meets the objections of network operators, because if a network operator transfers the decision, a fine cannot be based on the fact that the relevant content was unlawful and that it had to be blocked or removed.

However, the way in which the selected control model is designed is met with some criticism. The three most important reasons for criticism are the following.

2. Concentration on a few offences preferable

The NEA refers to 22 criminal offences, which, upon fulfilment, lead to an obligation to remove the content within 24 hours. The application of this will be very difficult because these offences have in common that they can only be understood and

⁴³ According to an early NEA draft a singular violation could suffice.

⁴⁴ Cf. The statement of the expert hearing before the Bundestag on June 19, 2017.

interpreted in light of freedom of expression. Often violations of the honor of a person occur in the context of a dispute.

According to the understanding of the German constitution, freedom of expression must be rescinded when a statement touches human dignity.⁴⁵ Otherwise, it is a matter of whether a defamation of the person is the focus. According to the Federal Constitutional Court and the Federal Court of Justice the mere violation of the honor of a person is not to be classified as an attack on human dignity.⁴⁶ If there is a factual reference, the Federal Constitutional Court requires a careful consideration.⁴⁷ Hereby, the court sets out strict standards.⁴⁸

For example, recently, a lawyer called a prosecutor “crushed, mischievous, and stupid”. The court considered that the lawyer was under the impression that his client was wrongly prosecuted.⁴⁹ A precise analysis of the situation in which the statement had been made was necessary. Thus, it was not considered as a mere defamation or abusive criticism (“Schmähkritik”).⁵⁰

For human dignity to be touched, it is necessary that the attacker is denied the right to live as an equal person in the community, and that he is treated as a subordinate being.⁵¹ This is typically the case with racist insults. For example, last year, an image went viral on which a soldier is shown with a machine gun and the following caption: “The fastest asylum procedure in Germany, rejects 1400 requests per minute.” People who had shared this in social networks were sentenced to financial penalties for incitement of hatred.

The same applies for a photo of a waste incinerator with the caption: “Even warm in winter. Can accommodate 100,000 people at the same time.” In German society, there is a great consensus that such content cannot be tolerated.

The boundaries between art and satire are difficult to pinpoint. A right-wing populist politician recently said, “political correctness must be sent to the rubbish heap of

⁴⁵ BVerfGE 75, 369 (380).

⁴⁶ BVerfG, NJW 2010, 2193 (2195); BGH, 1 StR 641/88 -, juris.

⁴⁷ BVerfG, NJW 2010, 2193 (2195); BGH, 1 StR 641/88 -, juris.

⁴⁸ BVerfGE 93, 266 (294).

⁴⁹ BVerfG NJW 2016, 2870.

⁵⁰ BVerfG NJW 2016, 2870 (2871).

⁵¹ BVerfG, NJW 2010, 2193 (2195).

history”, a satire magazine commented: “One has to agree on that with the Nazi slut!” Her own quotation had been satirically turned against her.

Whether a degrading utterance must be tolerated in the public discourse depends on its concrete context. The expression of an opinion is protected as a fundamental right, and this protection accounts for opinions that are well-balanced, polemic, provoking as well as repulsive.⁵²

Considering these principles, which are enshrined in German constitutional law, it is clear that a duty to promptly remove or block can only be considered in a few exceptional – more or less unambiguous – cases. The specification of “unlawful content” must therefore be ideally as narrow as possible,⁵³ which infers that the definition in the NEA must limit itself to only a handful of offences and not 22. Experience shows that courts know how to stop excessive persecution of the authorities.

It would, however, be advisable to limit the scope of relevant provisions of criminal law to such offences that are capable of denying persons their right to exist, or leading to public peace disturbance. The following offences of Sec 1 (3) NEA typically touch human dignity: incitement of hatred (Sec 130 (1), (2) of the German Criminal Code), depiction of violence (Sec 131), distribution, acquisition and possession of child pornography (Sec 184b), violation of intimate privacy by taking photographs (Sec 201a). Further, an accelerated blocking of content seems reasonable if the content is very likely to entail a further tangible risk of legal infringement or if it leads to significantly lowering the psychological barrier for violence against others through aggressive speech.⁵⁴

This is typically the case in public calls for criminal offences. On the Facebook page “Do not give up Germany” a video about a person from Africa was published, there was a comment below “anyone who sees this guy, hit him, a thick stone around the neck and from a bridge into the water”. Also, calls for lynch law belong in this category.

⁵² Grimm, NJW 1995, 1697.

⁵³ Keats Citron, „Extremist Speech and Compelled Conformity”, University of Maryland Francis King Carey School of Law Legal Studies Research Paper No. 2017-12, p. 7, 20.

⁵⁴ BVerfGE 124, 300 (332).

In this aspect, five offences of Sec 1 (3) NEA are relevant: preparation of a serious violent offence endangering the state (Sec 89a), encouraging the commission of a serious violent offence endangering the state (Sec 91), public incitement to crime (Sec 111), breach of the public peace by threatening to commit offences (Sec 126) and incitement of hatred (Sec 130 (3)).

Unfortunately, the legislator has not followed this recommendation for restraint, because it appears that quick deletion results are more important than balanced quality decisions.

3. Time pressure and high fines boost the risk of an overflowing blocking practice

For assessing the question of whether a specific content is (manifestly) unlawful, the subjective perception of the responsible employee will almost unavoidably be a decisive factor. Also, it already has proved difficult for social networks to differentiate between (criminal) utterances and journalistic reports about these utterances.⁵⁵ The respective employee will tend to avoid the difficulties of a complex decision-making or legal classification. With the risk of high fines in mind, the networks will most likely be more inclined to delete a post than to expose them to the risk of a penalty payment.⁵⁶

It should also be noted that sanctions can basically only be imposed if a network operator does *not* remove *unlawful* content.⁵⁷ The removal of a *lawful* content is highly unlikely to be punished. This also nourishes the fear of an overflowing blocking practice.

The danger of “overblocking” can indeed be countered with the establishment of the Regulated Self-Regulation, at least to a certain extent. But it is still unknown which of the networks will be involved in it and procedural rules for the self-regulation facility – the core element of self-regulation – remain to be set out.

⁵⁵ <http://www.spiegel.de/netzwelt/netzpolitik/heiko-maas-droht-facebook-wegen-hasskommentaren -a-1103167.html>.

⁵⁶ Keats Citron, „Extremist Speech and Compelled Conformity”, University of Maryland Francis King Carey School of Law Legal Studies Research Paper No. 2017-12, p. 24; see also: <https://www.bitkom.org/Presse/Presseinformation/Bitkom-zum-Gesetzentwurf-gegen-Hasskriminalitaet-in-sozialen-Netzwerken.html>.

⁵⁷ Ladeur, Medienkorrespondenz 15/2017, p. 5.

4. Insufficient specifications for the procedure of removing content

In the NEA, the participation of the persons concerned, above all the author of a content, is only insufficiently regulated. The procedure begins with the complainant filing a complaint against a content. In regard of the complaint, the NEA only requires the complaint itself to suffice for assessing the unlawfulness of the content. This is relatively vague and leaves much room for interpretation.⁵⁸ A list of necessary aspects that should be included in the complaint could have made it clearer.

The author of a content is informed about the decision of the network. Prior to the decision, the network *can* – not *must* – give the user the occasion to state a position, if the decision depends on the lack of truthfulness of a factual claim or depends recognizable on other factual circumstances.⁵⁹

If the opportunity for stating an opinion remains unused, or is not requested by the network, the social network can assume the credibility of the complaint and, where appropriate, remove the content. If, on the other hand, the user defends his content, the network must weigh the credibility of the conflicting claims. It could happen that in retrospect the network made a false decision in this weighing process. In this case, no fine must be imposed.

Only self-regulation bodies must set up a complaints office for users whose content has been wrongly removed.⁶⁰ The legislator assumes that in cases of unauthorized blocking, networks will quickly restore such content. However, a “put back-procedure”⁶¹ is also required if operators do not join the self-regulation. Proposals, which were aimed at that, were eventually not implemented into the NEA.

VI. Outlook

Hardly any recent German law has attracted so much attention on the international level as the NEA. The OSCE Representative on Media Freedom⁶² and the UN

⁵⁸ Ladeur, Medienkorrespondenz 15/2017, p. 6.

⁵⁹ Sec 3 (2) a NEA.

⁶⁰ BT-Drs. 18/13013, p. 23.

⁶¹ <http://www.taz.de/!5424947/> oder <http://tabea-roessner.de/2017/06/19/netzdg-grundlegende-neubewertung-erforderlich/>.

⁶² Holznagel, <http://www.osce.org/fom/333541>.

Special Rapporteur on Freedom of Expression⁶³ commissioned expert opinions on the draft law. They fear that there will be imitation effects, which could result in a restriction of freedom of expression. The Russian government has recently poured oil into the fire by submitting a draft law which is based on the German NEA. Whether these fears will be confirmed in the future is difficult to predict now.

The legal and political discussion here is still ongoing. Presumably, the Federal Constitutional Court will soon be called upon to discuss the constitutionality of the NEA. Several MPs have already pointed out this possibility during the hearing before committee and are hoping to receive advice from "Karlsruhe" (seat of the Federal Constitutional Court). It will also be interesting to see whether the law will be properly implemented at all, especially by the social networks. Some of those are still undecided whether to participate in the self-regulation. Since there is no specific definition of a systematic flaw in complaints management, it is unlikely that they will be fined for the time being. Much of this will depend on the execution regulations of the Ministry of Justice, which are to be announced this year.

⁶³ Kaye, <http://www.ohchr.org/Documents/Issues/Opinion/Legislation/OL-DEU-1-2017.pdf>.

Annex: Selected offences of Sec 1 (3) NEA

- **Sec 129a GCC “Formation of terrorist groups”**

Sec 129a covers the formation of and the participation in an organization whose aims or activities are directed at the commission of enumerated capital offences. On the other hand, Sec 129a (2) requires the formation of or participation in an organization whose purpose is the commission of other – less severe than those under Subs 1 – criminal offences and which also aims to intimidate the public in a considerable way, to unlawfully coerce a public authority or an international organization through the use of force or the threat thereof, or to significantly impair or destroy the fundamental structures of a state or an international organization.

- **Sec 91 GCC “Encouraging the commission of a serious violent offence endangering the state”**

Punishable is a person that hands out a guide about the performance of serious State-threatening acts of violence. The guide must be of such nature that upon reading the provided instruction the averagely informed and interested individual is able to carry out such act.⁶⁴

- **Sec 130 GCC “Incitement of hatred”**

Incitement of hatred refers to the hatred against a national, racial, religious group defined by their ethnic origins. It also refers to hatred against segments of the population or individual because of their belonging to one of the aforementioned groups.

According to Sec 130 (1) the incitement to hatred against groups of the population is punishable if it is suitable to disturb public peace. Whether a certain incitement is actually capable of disturbing the public peace is measured by the intensity of the attack and the susceptibility of the addressed people.⁶⁵ Especially adolescents are very susceptible to agitations against asylum seekers or foreigners in general.⁶⁶ It is not necessary that the incitement takes

⁶⁴ Von Heintschel-Heinegg, Beck-Commentary of the GCC, 33rd ed. 2016, Sec 91 pt. 4.

⁶⁵ Schäfer, Munich-Commentary of the GCC, 3rd ed. 2017, Sec 130 pt 24.

⁶⁶ Schäfer, Munich-Commentary of the GCC, 3rd ed. 2017, Sec 130 pt 24.

place in a public setting.⁶⁷ It is also not required that the group that is being discriminated against is aware of the attack.⁶⁸

Sec. 130 (2) relates to the dissemination of writings which incite hatred against certain groups of the population.

Of particular relevance are Subs 3 and 4. According to Subs 3, the person who publicly approves, denies or trivializes Nazi crimes and thereby disturbs the public peace shall be punished. According to Subs 4, publicly approving, denying or trivializing Nazi rule is punishable if it is capable of disturbing public peace. The requirements of Subs 4 can only be met if the statement refers to human rights violations characterizing the Nazi regime (not covering, for example, the approval of motorway construction during that time).⁶⁹

- **Sec 185 GCC “Insult”**

An insult is understood as an attack on the honor of another person by issuing one's own disregard or disrespect.⁷⁰

If someone is confronted with true facts that objectively diminish his honor Sec 185 does in principle not apply.⁷¹

With his statement, the offender must express his own disregard. It is therefore not sufficient if only messages of others are shared. It may, however, be the case that the nature of the transfer reflects his own disregard. In the case of an utterance in social networks, it is particularly important to check whether a separate statement is made by clicking on the “share”- or “like”-button.

- **Sec 186 GCC “Defamation”**

The defamation consists of the assertion or dissemination of unproved facts against a third party, which can cause disregard by others. As a further distinction to Sec 185 is to be emphasized that Sec 186 only covers the allegation or distribution of facts, value judgments do not fall under Sec 186. The delimitation can be difficult. It is not necessary that the fact is untrue – it is sufficient that the proof of truth is not given.

⁶⁷ Schäfer, Munich-Commentary of the GCC, 3rd ed. 2017, Sec 130 pt 25.

⁶⁸ Schäfer, Munich-Commentary of the GCC, 3rd ed. 2017, Sec 130 pt 25.

⁶⁹ Rackow, Beck-Commentary of the GCC, 33rd ed. 2016, Sec 130 pt 38.

⁷⁰ Regge/Pegel, Munich-Commentary of the GCC, 2nd ed. 2012, Sec 185 pt 8.

⁷¹ Valerius, Beck-Commentary of the GCC, 33rd ed. 2016, Sec 185, pt. 21.

- **Sec 187 GCC “Intentional defamation”**

The provision largely corresponds to Sec 186. The deciding difference being, that the alleged fact must be untrue.

Fausses nouvelles et autorité numérique

Marcello Vitali-Rosati

De fausses nouvelles, il y en a toujours eu. Qu'il s'agisse de canulars, calomnies, propagande, ignorance, tromperie, mensonge, croyance, notre vie sociale regorge d'informations, d'idées, d'affirmations qui *ne correspondent pas à la vérité*. Pensons à Platon et à son besoin de distinguer l'opinion (*doxa*) de la vérité: la doxa est toujours assujettie au risque du faux. On pourrait même aller jusqu'à affirmer que le langage lui-même n'existe que pour dire le faux: on dit ce qui n'est pas, car autrement on n'aurait pas besoin de le dire. La vérité se montre toute seule et on n'a recours au langage que pour la dissimuler.

Mais qu'est-ce qui rend alors si particulier ce phénomène actuel que l'on qualifie de "post-vérité" ? Quel est le rôle du web et des médias numériques dans cette apparente explosion des fausses nouvelles qui semble caractériser les dernières années ?

En 2016, en effet, le mot "post-truth" a été choisi comme "mot de l'année" par les Oxford Dictionaries, en raison de la fréquence élevée de son emploi. En particulier, la sortie du Royaume-Uni de l'Union européenne et l'élection de Donald Trump aux États-Unis ont été expliquées par plusieurs analystes comme le résultat d'une présence inédite de fausses nouvelles qui ont fortement contribué à manipuler l'opinion publique. Peu après l'élection de Trump, Kellyanne Conway, conseillère du président, utilise l'expression "alternative facts" pour défendre les fausses affirmations de la Maison-Blanche concernant le nombre de participants à la cérémonie d'investiture du président. Ces évènements qui semblent mettre en crise notre rapport à la vérité ont été souvent attribués à l'impact des médias sociaux qui permettent la circulation et la diffusion rapide d'informations non vérifiées. Pour résumer: les lecteurs sont bombardés de fausses informations sur le web, les réseaux sociaux en augmentent la visibilité si bien que les fausses nouvelles finissent par avoir plus de poids - et plus de crédibilité - que les vraies. Cela détermine une situation où il n'est plus possible de distinguer le vrai du faux, jusqu'à mettre en crise la notion même de vérité.

Le rôle du web

L'idée que les technologies numériques, et plus précisément le web, seraient les responsables d'une telle situation se base sur un présupposé qui me semble profondément faux: le web serait un espace sans règles ni structure où "n'importe qui peut dire n'importe quoi".

Cette idée n'est pas seulement une vulgata, mais elle est souvent présente aussi dans le discours d'intellectuels qui analysent les changements produits par le numérique. Que l'on pense à l'affirmation d'Umberto Eco, selon lequel «Les réseaux sociaux ont donné le droit de parole à des légions d'imbéciles qui avant ne pouvaient parler qu'au bar, après un verre de vin, sans conséquence pour la collectivité. On les faisait se taire tout de suite, tandis que maintenant ils ont le

même droit de parole qu'un Prix Nobel.» cf. ici. Outre que dénoter une certaine imbécillité de mon fameux concitoyen - même les érudits peuvent donc être imbéciles -, cette phrase nous révèle quelque chose d'essentiel que je vais essayer d'analyser en ces quelques pages: le rôle de l'institution dans l'établissement de la vérité.

En effet, plutôt que d'opposer un espace sans règles ni structure - le web - à un espace structuré et ordonné - l'espace pré-numérique -, il est nécessaire de comprendre que le web est lui aussi très rigidement organisé. Ce qui pose problème est que son organisation diffère de celle de l'espace pré-numérique, et en particulier de l'espace médiatique pré-numérique. En d'autres termes, il n'est absolument pas vrai que sur le web tout est sur le même plan et que c'est pour cette raison que l'on ne peut plus distinguer le vrai du faux. Au contraire, sur le web chaque objet, chaque texte, chaque document, chaque information occupent une position et une place bien précise et sont insérés dans une hiérarchie très structurée. Mais cette hiérarchie n'est pas celle de l'espace médiatique pré-numérique.

La question qu'il faut donc se poser est : quelles sont les conditions qui permettent à une information d'avoir plus de crédibilité qu'une autre?

Autorité, confiance et vérité

Cela revient à se poser la question de l'autorité. Et à y regarder de plus près, c'est justement cette question que soulève Umberto Eco avec toute la vulgarité de sa triste phrase. Commençons d'abord par donner une rapide définition de l'autorité. On peut se baser sur l'idée d'Hannah Arendt selon laquelle l'autorité est la capacité d'obtenir de l'obéissance sans avoir recours ni à la persuasion ni à la contrainte. En d'autres termes, l'autorité est quelque chose qui inspire confiance: nous croyons à l'autorité non parce qu'on nous démontre qu'elle dit le vrai, ni parce qu'on nous oblige à y croire; nous croyons à l'autorité parce que nous lui faisons confiance.

Pourquoi, selon Eco, le prix Nobel devrait dire le vrai plus qu'une personne qui boit un verre de vin au bar? Parce que le prix Nobel est une autorité. Nous faisons confiance au fait que quelqu'un qui a reçu un prix Nobel soit un expert, un savant et soit donc en mesure de nous dire la vérité. Dans toute société il y a une structuration précise et claire de l'autorité qui nous permet de savoir en quoi - et en qui - avoir confiance. On reconnaît toujours les signes de l'autorité. Il existe en effet une organisation d'institutions qui permettent de reconnaître : les États, les Universités, les médias, les maisons d'édition... La vérité n'est possible que grâce à cette organisation: les institutions garantissent des critères de vérité et ont l'autorité pour les faire respecter. Dans les journaux, par exemple, l'institutionnalisation des critères de vérité est le fruit d'une longue histoire et d'un ensemble de critères : les lois, les déontologies professionnelles, de longues négociations du rapport de confiance avec les lecteurs, une position particulière

dans une certaine société... permettent à un journal d'avoir d'une part des critères de vérité clairs et stables et d'autre part de gagner de l'autorité.

Autorités numériques

La rapide diffusion du web a quelque peu bouleversé ces institutions en produisant de nouveaux dispositifs d'autorité. Si l'on prend en considération de manière superficielle ce qui s'est produit, on est porté à croire qu'il n'y a plus d'autorité, que tout est sur le même plan. Mais à un regard plus attentif, il est facile de réaliser que c'est le contraire qui est vrai. Les plus grands acteurs du web sont justement ceux qui arrivent à produire confiance et autorité. On pourrait même dire que ces grandes entreprises vendent de l'autorité. Le web est une énorme machine de production de l'autorité. Pensons à l'autorité que nous accordons à *Google search*: 99% des usagers ne vont jamais au-delà de la première page de résultats. Cela signifie que nous considérons que *Google search* nous dit la vérité: ses premières réponses sont les bonnes. La confiance que nous accordons à de plateformes comme Facebook, Wikipédia, Amazon... nous montre qu'il n'est absolument pas vrai que tout est sur le même plan. Une information qui est présente sur les murs Facebook d'un million d'usagers n'est pas sur le même plan qu'une information qui n'est sur aucun de ces murs; un livre qui apparaît sur la page d'accueil d'Amazon pour un million d'usagers n'est pas sur le même plan qu'un livre qui n'y apparaît pas. Un blogue qui est listé en premier dans une recherche sur *Google search* a une position complètement différente par rapport à un blogue qui n'est pas indexé.

La question est de comprendre comment l'agencement institutionnel qui permettait la production de l'autorité dans le monde pré-numérique est en train d'être restructuré. Les équilibres changent et en effet, le fait d'avoir reçu un prix Nobel, d'avoir été publié par une grande maison d'édition et d'avoir fait la une d'un quotidien important ne sont plus les seuls paramètres pour acquérir de l'autorité. Il y a désormais d'autres dispositifs de production de l'autorité et la position que l'on occupe dans l'espace numérique en est l'un des plus importants. Cela ne veut pas dire que les autorités pré-numériques n'ont plus un rôle fondamental à jouer. Dans l'espace numérique, les autorités traditionnelles continuent d'exister : le site web d'un gouvernement a plus d'autorité que celui d'un groupe privé, l'affirmation d'un prix Nobel a plus de poids que celle d'un inconnu. Mais il y a désormais d'autres dispositifs qui produisent de la confiance et de l'autorité en parallèle - voire en concurrence - des anciens.

Que faire?

Au lieu de crier au scandale, les institutions doivent essayer de comprendre ces mécanismes et tenter d'en devenir les protagonistes. Il me semble que la question principale à se poser est celle de l'espace public. Un des problèmes fondamentaux des dispositifs de production de l'autorité sur le web est que la quasi-totalité

d'entre eux est privée. Faire confiance à quelqu'un parce qu'il détient un diplôme universitaire signifie faire confiance - au moins dans des États où l'université est publique - à une institution qui appartient à la collectivité. Ses choix et sa façon de déterminer des critères de vérité - par exemple des méthodologies de recherche - sont négociés de façon publique. Sur le web, à part quelques exceptions comme Wikipédia, l'autorité est concentrée dans les mains de quelques entreprises et ce sont ces entreprises qui ont le bénéfice d'établir des critères de vérité comme bon leur semble.

Lors de l'élection de Trump, Zuckenberg a essayé de mettre en place des systèmes pour limiter la circulation de fausses nouvelles sur Facebook: de cette manière, il revendique le rôle institutionnel de Facebook qui détient une forte autorité, et il affirme sa responsabilité dans la définition des critères de vérité. Je ne vois rien de mal, en soi, à cette situation. Un acteur privé qui a de l'autorité - comme peut l'être aussi un quotidien, par exemple - se trouve face à une crise de la véridicité des informations qu'il contribue à faire circuler - comme cela pouvait arriver aussi à un quotidien classique. Il décide donc d'agir pour rendre plus fiables ses critères de vérité - comme l'aurait fait un quotidien désireux de récupérer la confiance de ses lecteurs.

Le problème n'est donc pas tellement le fait que Facebook ait de l'autorité, mais plutôt qu'il y ait si peu d'autres institutions capables de produire de l'autorité en ligne - laquelle se retrouve, par conséquent, excessivement centralisée. Certains auteurs (comme Morozov ou Sadin) considèrent qu'il est impossible de contrer ce phénomène, mais il me semble que des expériences comme Wikipédia montrent le contraire. Wikipédia est parvenue à s'ériger en autorité tout en négociant collectivement, de façon ouverte et publique, ses critères de vérité. Et l'impact de cette expérience est tout à fait comparable à celui des grandes entreprises du web. Le succès de Wikipédia est justement basé sur sa capacité à mettre en place un dispositif d'évaluation et de vérification fortement structuré et adapté aux structures de l'espace numérique. Certes ce ne sont pas - ou pas principalement - des prix Nobel qui écrivent les articles, mais les dispositifs de validation des informations sont assez bien structurés et stabilisés pour que l'on puisse faire confiance à Wikipédia autant qu'à l'académie des Nobel.

Wikipédia reste un cas plutôt isolé. Mais le web pullule d'initiatives qui permettent ainsi de négocier des critères de vérité et de produire de l'information de qualité. Si nous voulons faire quelque chose pour contrer l'explosion des fausses nouvelles, nous devons d'abord comprendre les mécanismes de production de la confiance en ligne et ensuite essayer d'investir cet espace afin de produire des modèles différents de ceux des grandes entreprises du web. Les médias traditionnels peuvent et doivent le faire s'ils veulent survivre. Des initiatives comme le fact checking me semblent des plus heureuses: mettre à disposition des lecteurs les sources en utilisant des hyperliens est une pratique simple et adaptée à l'environnement numérique. Le plus de liens seront présents dans une information, le plus elle sera auto-vérifiable.

Toute attitude paternaliste et méprisante face à l'espace numérique me semble

destinée à l'échec - outre qu'au ridicule. Le web et les environnements numériques en général sont devenus désormais notre principal espace de vie: il est donc fondamental de créer des lieux où l'on puisse négocier de façon collective et publique les critères de vérité.