

Sabrina Slipchenko

Bard Center for the Study of Hate

Summer 2019

Berlin, Germany

Countering Hate at the Syrian Archive

I think that hate is a thing, a feeling, that can only exist where there is no understanding.

Tennessee Williams, foreword, *Sweet Bird of Youth*

“You’re going to make yourself sick with hatred.”

“No, I’m not. I’m going to make myself well.”

James Baldwin, *Blues for Mr. Charlie*

I. The Unknown Knowns: Initial Reflections on Hate

Where does hate live? How does it emerge? What is our place in it? In reflecting on these questions, memories spring to mind: memories of signs held by hopeful hands at various marches, signs that announce loudly: “love counters hate!” There are memories of different signs too, signs that flutter proudly the phrase “Hass schadet der Seele”¹ from the iron gates of Berlin’s evangelical churches. I think of ominous signs, like the trend of anti-pluralist populism advancing across the West again, or the tattooed symbols that are secret signs of the far-right. I think of nightmarish notions of hate: as a timeless phenomena that spans across the history of the

¹ “Hate harms the soul.”

world, as an energy drink fueling the faceless masses that crave death and the destruction of whatever liberal order.

I think of hate “on the ground”, glaring hate that perpetuates itself in war, straightforward hate that leads to the aerial shelling of bakeries and schools. I think of the indirect hate of indifference-- of looking away, choosing to forget. I think of sweet childish hate, such as “I hate this ice cream flavor!” or “I hate this song!”. I think of a verse that is often repeated in Bible school, ²לא תשנא את אחיך בלבבך, the ancient call against hatred whose echo has somehow reached us³.

I think of the light hypocrisies in general musings on hate. Hate is something “out there”, something practiced by others that “we” work against, something that can only happen at a certain place and time, by certain people in certain conditions. Hate is bad, and don’t do it! On the other side are those that assure us that we all find ourselves somewhere on the hate spectrum, since hate is a human phenomenon. However, it is also a genetic mutation. But really, it isn’t simple. Nonetheless, it has to be stopped. And on and on.

Unprompted, I don’t think about hate *that* often. I’m not a “hater”; I’m not part of any nationalist militant group⁴. Like most others, I’m an average person. I don’t practice outright hate (except for the occasional road rage slip while biking). I try to educate myself against racism, to fight against climate injustice, to advocate for marginalized groups. I try to become aware of (and undo) my own ingrained propensity to “other”. Sometimes I try to focus on “love” instead,

² “Do not hate your fellow in your heart.” Vayikra, 19:7.

³ Reminiscent of something Martin Luther King Jr. said: “It’s wrong to hate. It always has been wrong and it always will be wrong... It was wrong in 2000 B.C., and it’s wrong in 1954 A.D. It always has been wrong, and it always will be wrong.” in “Rediscovering Lost Values” (speech, Detroit, Michigan, 28 Feb. 1954), Stanford King Institute <https://kinginstitute.stanford.edu/king-papers/documents/rediscovering-lost-values-0>.

⁴ Whew!

that fabled golden antithesis of hate. Love in whatever form it appears. Sometimes love is the presence of my fellow marchers; sometimes love is a rallying call on Twitter by a trustworthy politician; sometimes love is a long-awaited policy change. However, as Hannah Ardent once wrote in a letter⁵ to James Baldwin, “In politics, love is a stranger, and when it intrudes upon it nothing is being achieved except hypocrisy... Hatred and love belong together, and they are both destructive; you can afford them only in the private and, as a people, only so long as you are not free.” Yet Baldwin, in his later writings, wasn’t quite as straightforward as Arendt suggested. He mostly exchanged the “gospel of love” for a murkier understanding:

I know what the world has done to my brother and how narrowly he has survived it. And I know, which is much worse, and this is the crime of which I accuse my country and my countrymen, and for which neither I nor time nor history will ever forgive them, that they have destroyed and are destroying hundreds of thousands of lives and do not know it and do not want to know it. One can be, indeed one must strive to become, tough and philosophical concerning destruction and death, for this is what most of mankind has been best at since we have heard of man. (But remember: most of mankind is not all of mankind. *But it is not permissible that the authors of devastation should also be innocent. It is the innocence which constitutes the crime.*)⁶ [emphasis added]

These reflections raise pertinent, if unanswerable, questions in regards to my work with the Syrian Archive. What is the place between love and understanding? What does a deeper understanding look like in real time, when the “enemies” are unknown, often unintentional (in

⁵ Hannah Arendt to James Baldwin. 21 Nov. 1962, in *Zeitschrift für Politisches Denken*. <http://www.hannaharendt.net/index.php/han/article/view/95/156>

⁶ Baldwin, James. *The Fire Next Time*. (London: Michael Joseph, 1963). 16-17.

that they harm without realizing), or altogether faceless? How can effective action be taken against such a complex problem?

II. Hate goes Online

The real world of fighting hate: we strap on our superhero battlesuits, equipped with shiny leggings, flashy armbands, and goggles. We march into the center of conflict and hand out roses. Everyone cheers and the villains have a change of heart. This is what I imagined fighting hate would look like.⁷ Instead, we spend most of the day poring over the computer (with intermittent popsicle breaks).

Though my coworkers are probably real superheroes by night, the day-job of fighting hate isn't always glamorous. This summer I've been working at the Syrian Archive, a Syrian-led initiative currently based in Berlin, with partner organizations that collect content from conflict zones in Yemen and Sudan. Our objective is to aid human rights advocates, including journalists and activists, in documenting human rights violations in conflict areas. We develop open-source tools and methodologies for "collecting, preserving, verifying and investigating visual documentation" of warzones. I like to think of it as amassing a collective memory bank. I think of "memory" in the most literal sense. In everyday life, something occurs to a subject; this data is stored by the subject, to be retrieved for some future use. As I understand it, Syrian Archive's "memory bank" functions similarly in that the subject's *own memories* are stored. There are thousands of hours of videos from the frontlines, videos (not of) but *by* those impacted by the conflict. As such, the traditional gaze is inverted. In the majority of cases, it is not a flown-in media team that captures evidence for a special in the nightly news. It is not a reporter from

⁷ Give or take the roses.

far-away climbing into a warzone to take a few photos. It is this “warzone” looking, confronting, showing itself. Memories from people on the ground emerge as *.mov* files. People upload what they see; this constant flow of visual cues is stored as memory.⁸

The truth of war is incomprehensible. There are many reasons for this, including misinformation, sheer volume, fatigue, distrust, and perpetual confusion. Indeed, much like actual memories, videos don’t always show the whole picture. Nonetheless, without the organization of this mass of memories, human rights advocates remain stuck in the realm of incomprehensibility. Collecting is an attempt to humanize victims, to prepare for post-conflict resolution, to “help societies understand the true human costs of war.”⁹ Credible, reliable, and transparent documentation is essential to the project of accumulating truth. In isolation, this content can lose much of its power. When it’s organized: patterns appear, proof emerges, and propaganda can’t survive.

Much of the content that we archive is sourced from social media platforms. People upload videos of a collapsed hospital, or an airstrike on a bakery, or the dropping of an illegal cluster munition. The platforms that host this content are used as tools, as intermediary archives that attempt to launch “memories” into the public sphere. This is happening against the backdrop of a newly emerging democratic order. In her book *Twitter and Tear Gas*, Zeynep Tufekci describes it as such:

Digital technologies have clearly transformed the landscape, seemingly to the benefit of political challengers. Rising in opposition to crumbling, stifling regimes that tried to control the public discourse, activists were able to overcome

⁸ This is a preliminary analysis on the politics of archiving. A more nuanced analysis is necessary, but outside the scope of this paper.

⁹ “About,” SyrianArchive.org. <https://syrianarchive.org/en/about>.

ensorship, coordinate protests, organize logistics, and spread humor and dissent with an ease that would have seemed miraculous to earlier generations. A popular Facebook page, created to decry the beating death of a young man by the Egyptian police, had been the forum for organizing the initial Tahrir uprising and had mustered hundreds of thousands of supporters. An Egyptian friend of mine would later joke that this must have been the first time in history when a person could actually join a revolution by clicking on “I’m Attending” in response to a Facebook e-vite.¹⁰

As the anecdote illustrates, social media platforms have become gatekeepers of the public sphere (thus they have the potential to be incredibly democratic). Revolutions are born from them, especially in places where traditional media outlets are co-opted by the government or oligarchs, or in parts of the world that the news-cycle has forgotten. Even after the first shots of the revolution have been fired, social media platforms¹¹ remain a vital reporting and documentation resource for communities.

With this critical backdrop in mind, we would assume that “curbing hate online” means empowering the very people that leverage the internet to fight hate¹². Much to my dismay, mainstream government objectives often hinder the fight against hate. Their motives are couched in paternalistic and punitive language bereft of nuance. Their fight is a single-minded one: western governments have consistently and blanketly condemned social media platforms for harvesting terrorism. The fight against terrorism is predominately a fight against Islamic

¹⁰ Zeynep Tufekci, *Twitter and Tear Gas: The Power and Fragility of Networked Protest*. (New Haven: Yale University Press, 2017). xxii.

¹¹ Hereafter dubbed “SMPs”.

¹² Hate in the form of unjust authoritarian regimes, for example.

extremism -- white nationalism is sidelined.¹³ To fight terrorism, governments put pressure on social media platforms to moderate their content *fast*. This means faster takedowns, and no incentive to err on the side of caution. As an example, current EU regulation (formed in response to the Christchurch massacre) explicitly pressures platforms to control content. Recent regulation attempts to implement “upload filters” and a “1 hour rule”¹⁴. In a recent White Paper, the UK has announced the creation of a governmental internet regulation agency. Its objective? To fight terrorism on social media platforms. In the US, a recent homeland security committee hearing brought in top executives from the three major platforms. The conversation centered around deep-fakes, Russian election interference, and terrorist content-- grossly bypassing any awareness of the revolutionary potential that has emerged through these sites, and urging stricter moderation.¹⁵ Recent attacks on Section 230 of the Communications Decency Act would make platforms liable for the content that’s uploaded onto their site, increasing their incentive to err on the side of overinclusion when taking down content.¹⁶

To takedown content fast, social media platforms create multi-stakeholder “hash” sharing databases. In this way, troublesome content cannot be uploaded on multiple sites. Hashes, or text-based video identifiers, are shared between platforms. According to recent numbers, the majority of this content is flagged by AI. Thus another key problem arises: there is almost no

¹³ Lee Fang and Leighton Akio Woodhouse, “How White Nationalism Became Normal Online”. *The Intercept*. 25 Aug. 2017. <https://theintercept.com/2017/08/25/video-how-white-nationalism-became-normal-online/> and Janet Reitman, “U.S Law Enforcement Failed to See the Threat of White Nationalism Online”. *New York Times*. 3 Nov. 2018. <https://www.nytimes.com/2018/11/03/magazine/FBI-charlottesville-white-nationalism-far-right.html>.

¹⁴ This law is an updated, and more extreme version of the 24 hour rule, which imposes heavy financial penalties on internet intermediaries for not taking down “harmful” content within 24 hours. More here: Nele Achten, *Lawfare*, “Social Media Content Moderation: The German Regulation Debate” (blog) 27 Dec. 2018. <https://www.lawfareblog.com/social-media-content-moderation-german-regulation-debate>.

¹⁵ Interestingly enough, only conservative lawmakers were concerned with over-moderation as a form of censorship.

¹⁶ GIFCT, a multi stakeholder hash sharing database spearheaded by Facebook, was created in response to this pressure. Unfortunately, this also causes mistakes multiply across platforms.

transparency in how AI “chooses” the content that gets flagged, in how the algorithms are programmed to understand extremism. This allows mistakes to multiply across platforms. What distinguishes a violent video, from a video documenting destruction? Only context can tell. Yet, in the current arrangement, context-aware moderation is being replaced by faceless programs, about the internal organization of which civil society is told nothing. It is important to note that social relations are embedded in technology;¹⁷ for the internet, this spells the reproduction of bias. Safiya Umoja Noble explains this in her *Algorithms of Oppression*:

Part of the challenge of understanding algorithmic oppression is to understand that mathematical formulations to drive automated decisions are made by human beings. While we often think of terms such as “big data” and “algorithms” as being benign, neutral, or objective, they are anything but. The people who make these decisions hold all types of values, many of which openly promote racism, sexism, and false notions of meritoricity, which is all well documented in studies of Silicon Valley corridors.¹⁸

Free speech violations are rife when platforms attempt to moderate content in this way. The government pressure induces censorship, as it’s nearly impossible to consider context under tight time constraints. Often content that is nowhere near in violation of the first amendment¹⁹ gets taken down. Typically users don’t know why their content has been taken down; they aren’t told what rule it violates. Users face harsh punitive measures (such as account and upload

¹⁷ As in, technology is not value-neutral. It is born from, and contains within it, the history (and biases) of a given social order. See Herbert Marcuse, “Some Social Implications of Modern Technology”. *Studies in Philosophy and Social Science* 9:3 (1941), 414-439.

<https://pdfs.semanticscholar.org/51b0/6070040048422d9d71b71c5a1228bbc702c2.pdf>.

¹⁸ Safiya Umoja Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism*. (New York: NYU Press, 2018). 1-2.

¹⁹ “Imminent harm” is one of the only justifiable restrictions on speech. This principle is interpreted narrowly, and most content doesn’t violate it. See Nadine Strossen, *Hate: Why We Should Resist it with Free Speech, Not Censorship* (Oxford: Oxford University Press, 2018).

blocks), and don't have recourse by way of a fair appeals process.²⁰ It's a Catch-22: the of the policies aimed at curbing hate actually exacerbate it, by censoring the very people that utilize SMPs to fight hate.²¹

In my research, I've found that civil society groups have responded to this challenge in different (yet harmonious) ways. Some focus their advocacy on free speech and an end to censorship, others make privacy and protection their primary rallying point. EDRi insists that social media platforms cannot become "privatized law enforcement". Governments and social media platforms form tight webs of collusion behind closed doors; this approach is fundamentally undemocratic and silences civil society actors. EFF has launched a website called "TOSsedout" that tracks people negatively impacted by biased takedowns. Article 19 has advocated for the creation of a social media platform, a multistakeholder body that would include civil society organizations and social media platforms, and decide on emblematic cases that could amend platform's terms of service. In this way SMPs could rely on an independent mechanism to push back against government pressure, and companies would be able to understand how their changes affect civil society organizations before they implement those changes.

But, no matter how you spin it, isn't it a bit problematic to advocate for leaving violent content up? Do we really want a feed full of beheading and exploitation? Why should SMPs be forced to leave this content up, when they have a first amendment right to remove it as they see fit? Isn't it more hateful to allow hate to linger in the public sphere? Nadine Strossen answers simply: "hate speech isn't harmless, but there are better ways to address it than censorship".

²⁰ More on the pitfalls of content moderation in Jillian C. York and Corynne Mcsherry, *Electronic Frontier Foundation: Deeplinks Blog*. "Content Moderation is Broken. Let us Count the Ways". 29 Apr. 2019.

²¹ ... or at least counter the effects of NGOs attempting to counter hate.

There will never be a one-size-fits-all definition for hate speech. Errors of judgement, including vagueness and overblown assumptions, will always overshadow any potential gains.

Counterspeech is a more effective remedy. As Monika Bickert has noted, counterspeech tactics such as “downranking” and linking white nationalists to reformist pages are policies that Facebook has utilized successfully. This method isn’t a “hands-off” approach to countering hate; it’s a radically effective one, aimed at healing root causes instead of overblown symptoms.

David Kaye’s *Speech Police* articulated another interesting angle to countering this problem, an angle that announces democratic possibility in a refreshing way. He claimed that some content has a value even if it’s too violent to leave up. As such, it should be shared with law enforcement *and* civil society organizations. The Syrian Archive has a unique position in this debate, since free speech/anti-censorship advocacy doesn’t directly address the potential that certain (even violent) content has for human rights defenders. Pushing explicitly for access, as well as standing behind free speech advocates, is a promising way forward.

III. Looking Forward: Conversations about Hate at the Syrian Archive

By hatred we refer to the “working definition” of hate studies: “the human capacity to define, and then dehumanize or demonize, an ‘other,’ and the processes that inform and give expression to, or can curtail, control, or combat, that capacity.”²²

Even though I believe in the work, I found myself fighting feelings of fatigue and restlessness. What’s the point of collecting these videos, in the long run, if the evidence is loud and clear and nothing changes? Is that where hate lives, in the subtle bias against reasonable proof of atrocities? If that is the type of hate that we’re fighting against, the hate born from

²² Kenneth S. Stern, “The Need for an Interdisciplinary Field of Hate Studies”. *Journal of Hate Studies* 3, no. 1 (2004), 11. <https://jhs.press.gonzaga.edu/articles/abstract/10.33972/jhs.18>.

fatigue, what can be done to increase our impact? In a recent conversation with my coworker, we spoke about this conundrum. Paraphrased, she described her approach in this way:

We're dealing with the effects of hate, the way that hate wreaks havoc. We're confronted with the banality of it. Individuals in the system try to destroy lives, building by building, so that the opposition loses all will. So that the opposition loses all will to fight. When we watch these videos, we don't feel the bigger picture. When the individual pilots perpetuate these bombings, they probably don't feel it either. But that's what it is. That's how far complicity spreads. Sometimes, it almost feels as if we're increasing the banality of it for ourselves by watching so many videos over and over. You get sort of numb to it. Another pile of rubble. Commonplace destruction. But the act of collecting, organizing, preserving, and publishing changes the content. It goes from banal, mass compassion-fatigue to information that has value and ideally an impact. The act itself, of organizing, undoes the narrative of the banality of commonplace evil. The information becomes impactful.

At the Syrian Archive, we deal with violent forms of hate. The hate that we encounter is visceral, state-level, torturous, painful, inhumane. It is the hate that fuels propaganda and war. It is the hate that blows homes to bits. Yet there is another side to hate: the hate of indifference, of banality, of incoherent government policies that haphazardly curtail attempts to curb hate. It can be exhausting. Nonetheless, imagine a world in which no one attempts to fight against hate, in which "no one cares, so I don't have to" is the primary motto. The Syrian Archive attempts to shape the direction in favor of a hateless world, by refusing to succumb to hate's weapon of mass destruction-- apathy.

Will we ever make it to a hateless world? And if not, why bother? I return to James

Baldwin for words for the way forward:

GOLDSTEIN: Do you have good fantasies about the future?

BALDWIN: I have good fantasies and bad fantasies.

GOLDSTEIN: What are some of the good ones?

BALDWIN: Oh, that I am working toward the New Jerusalem. That's true. I'm not joking. I won't live to see it but I do believe in it. I think we're going to be better than we are.²³

And it's this hope, that hope that things can be better than they are, that pushes us forward. So perhaps a remedy against hate is not love, but small hope.

²³ Baldwin, James. "The Last Interview." by Quincy Trope. *The Village Voice*. November 14, 1987.