Nisa Danitz

Ken Stern

Bard Center For The Study of Hate

20 September 2020

Bard Center For The Study of Hate Internship

The Bard Center for The Study of Hate was created with the intention of not only researching hate but also working to understand best practices in the fight against hate. The Center is now turning its attention to a relatively new platform where hate has taken root - the internet. In response to this challenge, the Center is hosting a workshop that will explore how the internet can be used to fight hate both online and "in real life". The workshop will be held at Bard and will bring together traditional NGOs and specialists who have focused their work specifically on the internet and social media. In recent years, hate groups have relentlessly used social media on the internet to spread their messages, to radicalize and recruit new members, and to harass their targets. This online recruitment has had serious effects on the real-world. Major displays of hate such as the Unite The Right Rally in Charlottesville in 2017 or The Christchurch Shooting in New Zealand in 2019 are the direct result of social media used both as a means of radicalization and as a community that acts as an echo chamber of hate.

The question then becomes, how can those same tools be used to fight hate? Social media has enormous organizational and educational powers. The purpose of this workshop is to explore through interdisciplinary conversations how the tools of social media can best be used in the fight against hate both online and in the real world.

Workshop Format

The workshop will start by establishing the context of the current crisis and the need for a response. The purpose of this section is to explore the definition of hate. Participants will share the strategies they use to define and identify hate. By establishing this foundation, the workshop can then move into specific practices that participants have found to be most effective.

The next section of the workshop will look at the current situation to gain an interdisciplinary understanding of how hate and the internet are interacting. This section of the workshop will focus on experts from various fields including professors, lawyers, and members of NGOs, who are thinking creatively about social media platforms and the fight against hate.

In the final section of the workshop, participants will engage in a deep dive into essential questions that come out of earlier portions of the workshop to determine best practices in the field and find new solutions and connections through different interactions of ideas and strategies. The closing of the workshop will be used as a jumping-off point to create a larger network of participants.

The interdisciplinary approach of this workshop will allow for the generation of new ideas that would not have otherwise been possible. This is one of the many reasons that the Center emphasizes in-person participation. This is the best way to foster creative thinking, by bringing together people who would not otherwise have met or had a conversation in the same space. The current date of the workshop is unknown due to the COVID-19 pandemic and the health and safety limitations that are in place at this time.

Experience This Summer

During the summer of 2020, the importance of studying the relationship between hate and the internet became even clearer. Due to COVID-19 restrictions, social media became an

essential means of organizing social justice events such as the Black Lives Matter protests. Human connection and the internet has never been more intertwined.

The original purpose of this internship was to gain a better understanding of what hate is and learn about ways to combat it. I contacted a number of leaders in this field to ask this question. However, I found that thinkers in the field of media seemed to avoid discussing the exact definition of hate. Rather than focusing on defining hate, their work focuses on recording actions that they view as unquestionably hateful and fighting those actions in systematic ways or through grassroots means. The nature of the online realm is to react to specific actions taken by others. Accordingly, this paper will not focus on the definition of hate; instead, I will focus on the strategies practitioners are using in the field today.

There are a few major ways that people are currently fighting hate online, including changing platform code, doxing people who spread online hate, grassroots attacks against hateful groups, recording hateful messages spread online, working to change the policies within companies, and working to change the laws governments have to regulate major social media companies. Over the course of the summer, I spoke to people from different disciplines who are working or thinking about the field of media. The work I did this summer will be carried on by another participant who will assist in formatting the workshop.

Conversations and Research

I started my research with Kate Klonick to examine this topic from a legal perspective. Klonick is a professor of law at St. John's University and an affiliate fellow at the Information Society Project. Her work focuses on the current type of regulations of social media companies such as Facebook and how to regulate hate without stepping on free speech issues. The internet is a special case that creates a particularly challenging legal environment as a result of section

230 of the Communications Decency Act and the cases behind it, which provides immunity to online platforms for harm caused by information that is posted by its users. This immunity from liability has left companies in the challenging position of weighing the right to free speech and the limitations of hate speech. Combating hate is further complicated by the fact that these social media companies act on a global level, interacting with different cultures where power structures differ. In the absence of a strong legal enforcement framework, much of the regulation of these companies is internal. Klonick's work focuses on how to govern these platforms. Other thinkers in the legal field are also worth exploring as this project progresses, particularly Danielle Citron, currently a Professor of Law at Boston University.

The Dangerous Speech Project (DSP) addresses dangerous speech rather than hate itself. When I spoke with Cathy Beurger, DSP's Director of Research, she emphasized that since hate is such a hard topic to define, DSP focuses more directly on how hate speech relates to the community affected rather than on the act of hate itself. DSP conducts research and provides services to mitigate the harms caused by dangerous speech in the future. The Dangerous Speech Project works on a global scale and they have a variety of projects that relate to the subject of dangerous speech in different countries. DSP recently published a paper called "But Facebook's Not A Country: How To Interpret Human Rights Law For Social Media Companies", which focuses on how challenging it is to regulate the internet because of its international base of users. Rather than focusing on individual acts, this paper discusses the framework that has allowed for abuses to take place.

Unlike The Dangerous Speech Project, Andrew Marantz had a very clear understanding of the mentality of people he followed for his book *Antisocial: Online Extremists, Techno-Utopians, and the Hijacking of the American Conversation.* Marantz, a writer for the New

Yorker, is clear about his ethical stance and its relationship to the opinions of those that he followed and wrote about in his piece. Marantz's piece followed and observed members of the alt-light. During our telephone conversation, Marantz emphasized the importance of taking a stance on hate when observing hateful group members. Other groups have taken a variety of different stances on the relationship between hate and the internet, most often focused around the idea that it arises in small, tight-knit communities that are generated using social media and its algorithms. Some argue that it is a matter of focusing on the "real world", and, contrary to Marantz, argue that online hate should not be reported on whatsoever because it amplifies voices that otherwise would be limited to a particular online space. Still, other groups, such as Global Voices take a different view by offering multiple perspectives and providing an outlet as well as training and media tools so that local communities can tell their own stories. Global Voices works globally through a volunteer network with a particular focus on human rights and empowerment and offers a wide selection of readings about social media companies beyond the United States. This form of story-telling and record-keeping is an important resource in fighting hate online. The Advox team, which empowers local communities, is particularly relevant to this workshop.

Nathan Matias of the Citizen and Technology Lab focuses his work on individual communities. When I spoke to Matias, he emphasized the importance of communities online and the harm that they face. His work has particularly focused on creating more inclusive spaces online. Rather than seeing hate as a generalized problem that affects all people in the same way or trying to determine what can be solved by "a button or single line of code" he views his work through the lens of the affected community. This is particularly relevant on platforms like Reddit where minority groups are often attacked. Matias takes a preventive approach by adjusting

messaging to deter hate. Other people that are thinking in a similar vein of online action include Joan Donovan, Adrienne Massanari, Betsy Levy Paluck, and Graphika.

Social Media companies have an interest in having their users stay on their platform longer. This is often done by pushing content that is more and more extreme and speaks to their biases. Different social media companies who have been involved in the promotion of hate speech in the past have started to intervene. One of the most receptive companies to proactive change thus far has been Reddit. This is in part due to Reddit's set-up, which is already monitored by the heads of each individual online community. Before this summer, Twitter seemed vehemently against any form of regulation of speech. Their commitment to free speech was clear and uncompromised by the hate and misinformation that was being spread on the application. During the COVID-19 pandemic, Twitter began warning users about misinformation. Although hate is still widespread on the application, this is an important step in reducing the spread of misinformation and openness to regulating what people say on their platform. All this demonstrates that different applications take different amounts of responsibility for the messages people spread on their platforms. As federal regulations have been so relaxed, and Section 230(c) protects online platforms from liability, the social media companies choose how much responsibility they will assume for what is posted on their platforms. Facebook, for example, has been in the spotlight of this push and pull between its various roles as a regulatory body, an advertiser, and as a forum for free expression. All of these roles are in tension and often contradictory, and the goals do not inherently align, causing Facebook, and other social media platforms, to have to make a decision as to what should be prioritized. Thus, much of what can be implemented is being determined by the companies themselves; often as a result of financial pressure and public relations. One of the most effective campaigns of the summer was a

partnership among Facebook's advertisers and traditional NGOs such as the Anti Defamation League (ADL) and the National Association for the Advancement of Colored People (NAACP) which called upon major corporations to remove their advertisements from Facebook until it took steps to remove hateful content from the platform. The campaign has been somewhat successful as Facebook has begun implementing important changes. Among the first steps to reduce hate is for social media companies to begin adopting internal changes in order to ultimately see what is most effective in reducing hate online since so much is currently unknown.

Issues In The Field

One of the largest problems that I heard consistently both in my conversations and through articles, is that it is extremely challenging to create legitimate change online because of the power that large social media companies have and the current legal framework that protects so much of online content. The issue of online hate is complex and delicate and is not fully addressed by the current legal framework. In the absence of effective legislation, organizations, and individuals who are attempting to address these issues often express frustration that they are not getting the opportunity to implement their innovative ideas about hate and the internet.

Takeaways and Future Progress

The study of online hate is currently exploding, as the historic moment, we are in relies so heavily on the internet. This has exposed the internet as a reflection and amplification of the real world, rather than a place that is immune to society’s shortcomings. Hate is defined in many different ways, often based upon the solution that is being worked toward rather than what causes the hate. This is most likely the result of the various disciplines that are tackling this intersection of social media and hate. Hate is still very much seen by practitioners in the field as something indefinable, but each industry has its own conception of identifying factors to indicate

that hate might be present. In Andrew Marantz's case, it is a dogma of hate, for Nathan J. Matias, it is direct requests from harmed members of the community. Organizations like the ADL which focus generally on larger patterns of hate have used their reputation to put financial pressures on social media companies. Kate Klonick and those who are a part of these larger social media companies try to find a balance between the American ideals of free speech and the protection of people on a global scale. These different philosophies all result in different calls to action. When these issues center around hate at the workshop rather than their individual practices, I believe the conversation around these issues could change. Rather than getting stuck on the differences in approach to hate, education, and interaction among the participants might deepen the conversation and lead to new ideas about what hate is and how best to combat it online.

Among the best practices in the field, a concentration on community seemed most important. Rather than getting stuck on the magnitude of social media, it is vital to focus on the community. Social media websites operate through the concept of community while approaching issues with a broad base solution that ultimately backfires. Working within communities and creating a clear understanding between members of the community and the social media platforms with clear expectations for users and the communities that exist within each social media platform. Through action-based in community and communication, these organizations are able to have a real impact rather than make changes that limit speech with no particular benefit. Another practice that has proven effective is to follow the money. Like the major campaign Stop Hate For Profit which was dependent on advertisement money. Another aspect of profit for company is the users themselves. As users of applications choose to prioritize hate in there choices on social media it incentivizes companies to regulate it more heavily and provide

better tools fight online hate. All of these approaches in the field require an understanding of both technology, sociology and current events.

Throughout my summer I was able to identify the primary players in the field of media and hate. Although I faced challenges this summer due to the unusual circumstances surrounding COVID-19, I was able to get diverse perspectives by reaching out to people online and on video calls. It was much more challenging to have in-depth conversations with people about the definition of hate. Much of what I was able to accomplish this summer felt like the beginning of a larger project. The next steps, beyond continued research, require focusing on how to format the event to build an environment for a creative and collaborative workshop with these remarkable innovators.

Possible new directions to take the project include the relationship between Artificial Intelligence (AI) and hate. Oftentimes AI has been used to generate hateful messages and propel them forward, but some people in the world of human rights and hate are working to find creative applications for a more positive purpose. Any conversation about AI will need to do a deep dive into the actual technology behind the issues. In this process, it continues to be important to discuss the idea of hate and the general state of the online community through legal experts, government officials, and representatives from large social media companies. The ultimate goal of the workshop to emphasize when approaching participants and organizing sections is how to encourage a space for conversation that would not otherwise be possible.