Oak Ritchie. Platform Algorithms and Their Effect on Civic and Political Arenas. A Master's Paper for the M.S. in IS degree. April, 2015. 66 pages. Advisor: Dr. Zeynep Tufekci

The purpose of this research was to analyze the ways that social media platform algorithms affect user experience and thus, civic and political arenas. This paper analyzes literature on the topic that will illuminate specific ways that algorithms dictate to the user what is visible or invisible, considered "fresh" or "irrelevant" and the implications of such digital curating.

Headings:

Social Media Platforms

Platform Algorithms

Political Sphere

Public Sphere

Algorithmic Curating

Facebook

Twitter

PLATFORM ALGORITHMS AND THEIR EFFECT ON CIVIC AND POLITICAL ARENAS

by
Oak Ritchie

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Information Science.

Chapel Hill, North Carolina

April 2015

Approved by

_____

Dr. Zeynep Tufekci

Table of Contents

INTRODUCTION

With the advent of the Internet came the ability to transmit messages around the globe, collaborate remotely, and share experiences and perspectives. This would have been considered a rarity, if not impossible, without such digital connectivity, and at present access to it remains reserved for the most powerful and wealthy. The scope of the technological structures for communication, collaboration, commerce, and entertainment that are being erected in this new landscape and how these spaces govern themselves bring challenges that require reflective thought and strong methodologies to solve effectively. Among the challenges for consideration is the human experience in digital spaces, and what tools will enhance and improve conditions for collective action.

The advancements in the iterative relationship between technology and the users of technology have produced complex systems that have become an inextricable part of the landscape, from the expectation of a grandparent to be able to place a video call with their grandchild to impatience when sending a message that takes longer than a few seconds. Human interactions have increased in online spaces, and status updates and posting of pictures and other content are increasingly becoming a form of engagement for those who aren't geographically close to those they desire to stay connected to. These online spaces, and the

mechanisms developed to construct them, bring with them benefits as well as challenges.

A twist to online spaces, such as social media, communications, entertainment and education are often spaces in which algorithms are making decisions for us, such as “Based on your watching this movie, you might like…” or “Other users who bought this product also bought…” Sometimes, that voice feels prescient in its ability to evaluate interest about some elements of human behavior, and machines are increasingly used to make such predictions about us.

My research question investigated the interaction between platform affordances; through which users can signal their preferences and interests, platform algorithms; through which content is made more visible or less visible; and how these policies and design choices impact political and civic arenas.

I. Platforms

1.1 A Brief History of Online Platforms

The World Wide Web, which is an information system comprised of connected hyperlinks that are accessed through the Internet, began in 1992 and was primarily a fairly limited space that institutions, like universities and governments, could use to transmit information back and forth as well as "browse" or "surf" using the hyperlinks and URLs available to them. The web, at that point had limited practical usefulness for an average user without some understanding of computing systems. By 2004, Web 2.0 had emerged with the capacity for users to interact with websites and other users in a more dynamic way. Instant messaging, e-commerce, search engines, and social media platforms grew with this development, and within a few short years the web became a public sphere where a user could generate, publish, share, and discuss content.

The advent of Web 2.0 brought with it access to a global public sphere where users have a basic understanding of the tools to browse and search could now find content of interest, communities with similar interests or ideals, and share content in a way that was previously limited by physical location. Communities grew in various spaces, and some users built spaces for specific communities to operate exclusively in political or non-political ways. Some non-political commercial platforms like LiveJournal, Friendster, MySpace and Facebook evolved and became

widely used to share content and interact in ways the platform made possible. Within these online spaces, there are users who are specifically motivated to enact change, be it political or non-political, and there are many who seek entertainment, social interaction, and comfort through funny memes, cute cat videos, or both (Zuckerman, 2014).

1.2 Current Platforms

A platform, in the broadest computing sense, is a framework upon which applications can run. A computing platform can be thought of as the environment in which the operating system, runtime libraries, and hardware architecture operate. In the context of the web, the framework allows for all manner of programs to be run within a space that is specified by a Uniform Resource Locator (URL), which is a unique location within the web that can be accessed by anyone who has the URL and is given access to its contents. On the web, the limits of what applications can link to outside of that platform are constraints of the platform itself. In the case of the Facebook platform, which was launched in May of 2007, an Application Program Interface (API) was made available to the public to create applications that fit its specifications to run "apps" that ran games, donate to other FB users, and send gifts among a wide berth of possibilities for developers. The data flows through that platform and user profiles operate within it.

The user creates a "profile" in this platform, which can be viewed as an online version of his or herself, to be broadcast in a public space for any with access to that profile to view. Depending on the purpose and privacy controls for any given

platform, this could be more secure or less secure for the user. Some platforms provide *user controls*, or ways for the user to make choices about the settings of their account, as a part of the *platform affordances* offered.

1.3 Platform Affordances

In the context of the field of Human-Computer Interaction (HCI), which studies how humans use computing systems, artifacts and infrastructures, an *affordance* is a discoverable action that a user can engage. Ideally, affordances on whatever platform the user is engaging will be somewhat intuitive, at least the basic conventions of user controls that are familiar to the user. For instance, if an experienced user of websites comes across a blank screen with one text word or a logo, they will instinctively attempt to engage that convention of browsing websites, which has come to imply the "Home" button. In order to clarify the concept of platform affordances, the following section describes user controls for two commercial platforms, which are the platform affordances that individual users have the ability to access and manipulate from their point of access on the platform.

1.3.1 User Controls

The architecture and user controls that a platform offers shape what the user can effectively do on any given network. The buttons chosen, the freedom to customize, a transparent policy on data gathered and how private, birthday notifications, and various other options are possible for the designers of a platform to adopt to define the user space. What user controls are available, as well as those that are not, inform us as to the orientation or the vision of the creators of these

spaces. The values of the creators and designers, whatever they may be, can't be divorced from the commercial products that they produce and maintain.

On Facebook, the primary user controls are the "Update Status" bar, the "Like" button, the "Comment" button, and the "Share" button (See Fig. 1). The "Like" (#1) button is a control to indicate some form of interest from one user to anyone with access to the network.

The "comment" (#2.) feature enables any user with authority to post to a comment thread on another user's feed. This feature allows for content to be generated whether it comes in the form of a conversation, sharing of memes or pictures, or a random thread of comments on a post on this platform. Through it, one connection in a network can become acquainted with a new pool of connections in the user network being visited as well as simply contribute content. The "share" (#3.) feature allows a user to share posts from their own page to another user, share another post on their own or another News Feed.

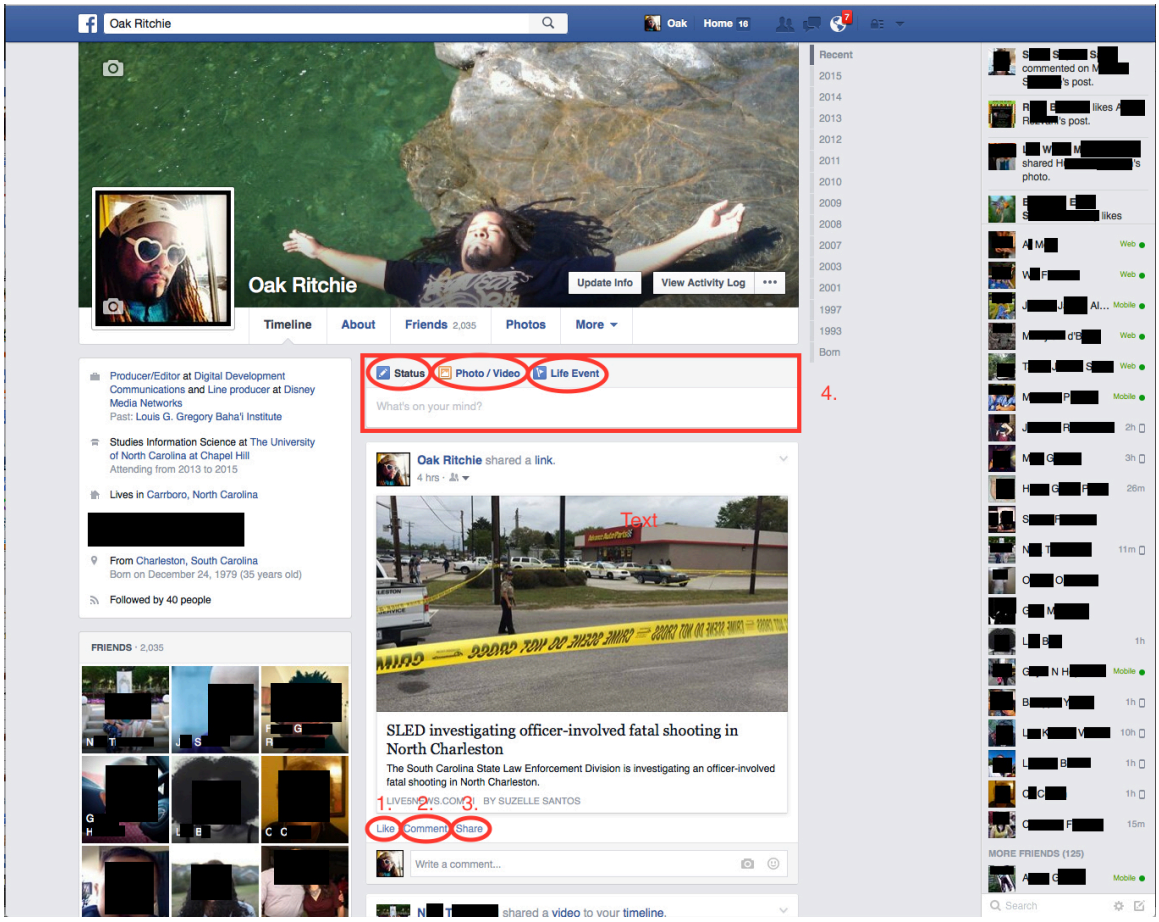

Fig. 1 Shows a typical 2015 Home Page, with controls as they were at that time.

As of this writing, on the desktop version of the site, the "status bar" (#4.) prompts the user to share with their network with this prompt: "What's on your mind?" This bar gives the user the option to post a status update, post a photo or video and post a life event, which then offers a drop-down menu (See fig. 2) to choose between "Work & Education", "Family & Relationships", "Home & Living", "Health & Wellness" and "Travel & Experiences". Upon choosing "Work & Education", a drop down menu gives the user the choices: "New Job", "Published Book or Paper", "Retirement", "New School", "Study Abroad", "Volunteer Work", "Military Service" and an option to "Create Your Own" event. The "Family &

Relationships" option (See Fig. 2) prompts the user to share "First Met", "New Relationship", "Engagement", "Marriage", "Anniversary", "Expecting a Baby", "New Child", "New Family Member", "New Pet", "End of Relationship", "Loss of a Loved One" and "Create Your Own" as options. The "Home & Living", the user can choose "Moved", "Bought a Home", "Home Improvement", "New Roommate", "New Vehicle" and "Create Your Own", if one so chose. The "Health & Wellness" option prompts the user to share if they became an "Organ Donor", "Overcame an Illness", "Quit a Habit", "New Eating Habits", "Weight Loss", "Glasses, Contacts, Other", "Broken Bone", "Removed Braces" and "Create Your Own" as options.

Fig. 2 The "life event" menu and options for sharing publicly or within network.

Each of the above options pulls up a form for each event, with options to take pictures of some things, enter dates and times of others, giving the user a rather exhaustive ability to publicly share details of their life. The user-generated turn of

the web consists of users "volunteering" their information to the networks they want access to and, in so doing, giving significant amounts of granular data to the provider (Gillespie, 2015).

On either side of a user's feed, there is an array of content displayed according to user preferences (See fig. 3) ranging from "Pages", "Groups", "Apps", and so on. To a large extent, this content can be edited. On the far right, there is a stream of people in the network of the user prioritized from top to bottom by those who are online presently down to how many minutes it has been since a connection logged off of Facebook. Just above that in the same column is a running ticker of "Likes", "Comments" and "Posts" that connections are making in real time.

Fig. 3 Sidebars, network feed and controls.

Finally, the top bar of the user interface has 7 controls: The first displays the name of the user with a profile picture next to it, which will take the user to their home page with content that they posted most recently. The second will take them to the News Feed page. The third will display “friend requests”, indicating other users that wish to connect. The fourth is the messenger, where personal messages from other users are accessible. The fifth is for activity on the user’s page, such as “comments”, “likes” or any other user interaction with the home page. The sixth produces a drop down menu that displays the privacy setting options that the user has available to choose from. The last button allows the user to use Facebook as an administrator for any page that they are managing, as well as the following options: “Create Page”, “Manage Pages”, “Create Ads”, “Advertising on Facebook”, “Activity Log”, “News Feed Preferences”, “Settings”, “Log Out”, “Help” and “Report a Problem.”

The controls on Twitter offer a smaller number of options, compared to Facebook, that are displayed almost entirely on the home page of the user profile (See Fig. 4). The desktop version of the Twitter “Home Page” displays in the center column “Tweets”, or messages from users to which the account holder is connected. These tweets are displayed as a chronological live stream with the most recent content displayed at the top. Above the user feed is a set of tabs, entitled: “Tweets”, “Following”, “Followers”, “Favorites” and “Lists”. Also included in the profile page of a user are two widgets, titled “Who to follow” and “Trends.” The “Who to Follow Widget” suggests users that might be of interest based on associations the algorithm has made with commonalities in user behavior, interests or activity. The last widget

contains “Trends”, which is a collection of popular hashtags and topics, indicated by the pound symbol (#) and followed by a stream of characters that have been used by users to describe a common event, person, idea or thing that is popular on the network at that time. The use of the “hashtag” in this way is something of a Twitter convention that users began employing, and the practice spread beyond that platform. Finally, there is a bar at the top of the interface with a “Home” button, “Notifications” button, “Messages” button, “#Discover” button, the Twitter logo, which upon selecting will refresh the Feed, a search bar, a “Profile and Settings” button, indicated by the user profile picture, the “Favorite” button and finally, the “Tweet” button, which allows the user to post to their feed.

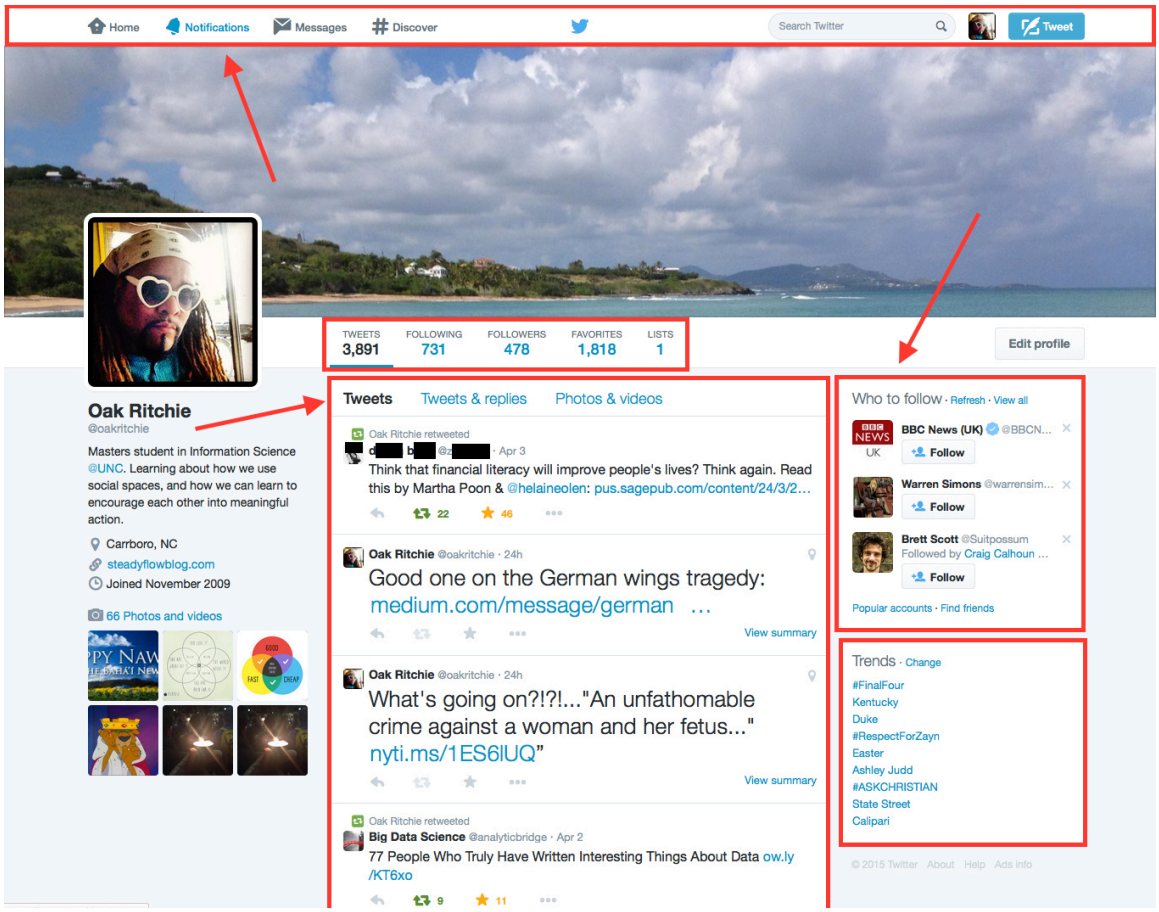

Fig. 4 Twitter User Profile Controls

The above examples are just two current designs that exist among many. How each platform is designed will dictate the usefulness of that platform for any specific task. Twitter, for instance, would have limited effectiveness in being used as a place to shop in comparison to a platform like Amazon.com or Ebay.com, which were constructed with finding and purchasing goods as the primary purpose of those platforms. Each user affordance is a prompt of what the designer intends for the user to accomplish on any platform, be it “Broadcast Yourself” (YouTube), “What’s on Your Mind?” (Facebook), or “What’s Happening?” (Twitter).

II. Algorithms

2.1 Define an Algorithm

Algorithms, according to one definition, can be thought of as "encoded procedures for transforming input data into a desired output, based on specific calculations" (Gillespie, 2014, p. 1). An algorithm can be coded to perform calculations, data processing, and automated reasoning among other things. The range of the scope of the design of an algorithm is virtually limitless, only to be subject to the constraints of its creation and production: budget, vision, and expected deliverable date being a few common constraints. They can be designed to automate simple tasks like reporting and regulating the temperature in a room, or complex tasks like filtering the flow of information within a network of billions of users, as is the case with commercial online platforms.

Some of the more complex algorithms are an outpouring of a field at the intersection of Computer Science and Statistics (and closely related to the study of human and animal behavior in Psychology) called "Machine Learning," which specifically explores how to get computers to program themselves. For example, an iPhone out of the box has only its programming, but the learning algorithm within it allows it to "learn" user behavior and predict things such as what word the user intends to type based on the input behavior of the user.

Nicholas Diakopoulos (2014) categorized the specific technical decisions that algorithms make, and broke them down into four types: prioritization, classification, association and filtering. While any algorithm can be designed to make any combination of these decisions, it is useful to consider the decisions separately. Prioritization describes the process by which data is included or excluded to get the results desired by the creator. What data is included or excluded for prioritization may make processes more efficient, more complete, and more detailed, depending on the needs of the intended design (Diakopoulos, 2014).

Classification involves the algorithm breaking each entity down into classes that it will fit the criteria for. A simple example of this would be an algorithm designed to separate a collection of text in to “positive” or “negative” terms. In the case of a list of movie reviews, a classification algorithm may be tasked with identifying which reviews are likely to be “positive” or “negative” based on the training data that it was given. A shortcoming of classification algorithms is that they produce a “false positive” or a “false negative”. A “false negative” is when the movie review had words that the algorithm was trained to see as “negative”, but some nuance of the use of words in the review made it into a “positive” review. A false positive is when the algorithm was trained to see the results as “positive”, but the review turned out to be negative (Diakopoulos, 2014).

Association deals with data that, because of the input of association between entities, whether words or actions, they are considered linked in some fashion. Another example can be recommendations for other purchases on Amazon.com. Upon browsing, creating a “wish list” or purchasing an item, the user is prompted to

consider items that their algorithm calculated were linked. One example of this was in 2012 when a group of scientists working for Target managed to design a system that could make associations from ordinary user purchases to predict certain changes in a user's life. An angry father entered the store, furious that his High School aged daughter received coupons addressed to her for baby clothes and cribs from Target. A few days later, the father had apologized because his daughter was indeed pregnant. Associations between seemingly innocuous purchases, changes in schedule, or some other behavior the designers had considered to be predictors accurately identified a user to be pregnant before family members were aware of it (Duhigg, 2012).

One definition for filtering describes it in this way:

> "**Filtering** involves including or excluding information according to various rules or criteria. Inputs to filtering algorithms often take prioritizing, classification, or association decisions into account. In news personalization apps like Flipboard, news is filtered in and out according to how that news has been categorized, associated to the person's interests, and prioritized for that person. Filtering decisions exert their power by either over-emphasizing or censoring certain information" (Diakopoulos, 2014).

These types of decisions that systems are programmed to make aren't exclusive to each other. There are algorithms that make all four kinds of decisions for the user regarding what content the user "would be most likely to want to see" in the estimation of the program. How much control and awareness the user has about these decisions made not entirely with their consent is something to consider.

The algorithms used by an organization or platform will at times be publicly available, as is the case with Reddit and Hacker News. Conversely, as is the case with Facebook and Google, the algorithm is behind what is called a “black box”, which simply indicates the contents of some part or all of the algorithm are not visible. (Pasquale, 2015) Three clear reasons exist that some platforms and organizations claim when defending their decision not to make their algorithms transparent, or publicly available. One is security, which in the case of governmental institutions is a high priority. The second is competitive edge, which takes the stance that the success of the platform is due to the unique qualities of their design, which would no longer be competitive if made publicly available. The third is to protect from “gaming” of the system, in which users who know what an algorithm rewards or punishes in “attention”, traffic or exposure exploit that knowledge to the extent that the system can even become ineffective (Gillespie, 2014).

Algorithms can be constructed in any number of ways, and it is important to consider an algorithm’s coding in relation to the calculations it makes, the scope of users and how it might affect the user’s perception. For instance, an algorithm designed to predict some aspect of human behavior cannot possibly have the entirety of the human experience inputted into it, nor could it presently be practical to attempt to do so. If the predictions are based on data from one subset of all of humanity, then the resulting output will predict based on that single subset. Whatever percentage, in all the ways possible that humans can be different, that this data excludes makes predictions and assumptions as though the excluded population does not exist. If this algorithm is intended to serve a local business with

marketing and advertising, the effect of this exclusion is limited. If it is intended to serve a network of billions of users and determine what they deem important, can share readily or even be aware of, then the effect of any exclusion has manifold implications (Gillespie, 2014).

The algorithm can be designed to predict detailed and granular information about users depending on the data collected about them. This prediction is only as accurate as the data and programming permit, and there is almost always "a margin of error" for tools of this sort. In addition to the limitations of scope of input data and the decision to include and exclude certain populations in the data set, there is also this margin of error already present in the algorithm's predictive capacity.

2.2 Applications of Algorithms

The application of algorithms can be seen in any virtual or physical space as well as any object that has a need for automation, calculation, or analysis for processes that have a desired outcome for the creator. They can be underlying the processes of a simple calculator, a system analyzing large streams of data, or a program designed to compose collections of texts that are very difficult to discern from a human-generated work.

2.2.1 Narrative Science: A Case Study

In the fall of 2014, the Associated Press automated their system to publish 3,000 stories that a human hand did not interfere with directly. The company that designed the AP's platform, Automated Insights, also has partnered with corporations such as Yahoo, Allstate and Comcast, generating millions of articles per week (Miller, 2015). One study conducted by Clerwall (2015) found that

participants saw a negligible difference between human-generated and software-generated journalistic content:

> "…the readers are not able to discern automated content from content written by a human. Some aspects of quality, such as being clear and being pleasant to read, received a slightly higher score for human-written content, but others, such as trustworthiness, informative, and objective, were higher for the automated content" (p. 11).

As was described earlier in this section, an algorithm requires data input to process and then output what the designer intends. In order for a system designed to reliably output sports reporting information, the scope of the data inputted is narrowed to include only useful information for that task. Steven Levy (2012) describes the steps that an algorithm designed to output sports reporting by a company called Narrative Science underwent. He shared that first, "it must amass high-quality data" and that then "the algorithms must fit that data into some broader understanding of the subject matter. (For instance, they must know that the team with the highest number of "runs" is declared the winner of a baseball game.) So Narrative Science's engineers program a set of rules that govern each subject, be it corporate earnings or a sporting event. The company has hired a team of "meta-writers," trained journalists who have built a set of templates. They work with the engineers to coach the computers to identify various "angles" from the data. Who won the game? Was it a come-from-behind victory or a blowout? Did one player have a fantastic day at the plate? The algorithm considers context and information from other databases as well: Did a losing streak end?" At that point, Levy describes

the adaptation of structure of writing and in order to construct sentences, “the algorithms use vocabulary compiled by the meta-writers. The company calls its finished product ‘the narrative.’” (Levy, 2012.)

During testing, this algorithm underwent a number of needed adjustments to avoid errors, not because of a flaw of the data inputted into the system, but rather in the emphasis and sentiment in the output from it to the audience it was designed to “report to.” If one team scores more points than the other, thus winning the game, doesn’t mean that the other team didn’t try hard or weren’t talented opponents. The input data early on in such algorithms had no concept of “sportsmanship” programmed into them, thus the binary “winner/loser” report would be made, which had high potential to humiliate further the team that lost. In other cases, all errors that children made in a little-league game according to the data were reported, which didn’t meet the desired results with the parental audience.

Even engaging a relatively small phenomenon in the human experience, such as sports, the lack of nuanced capacity to recognize that social capital was also at stake in addition to the benefits to winning and drawbacks to losing. A human news reporter is often trained to recognize and respond to the sentiments that are potential results of their choice of words. If the losing team made a valiant effort, they are aware that the audience would find value in some emphasis in that observation. If child players made errors or dropped a fly ball, being aware of the audience for such reporting would likely focus on the triumphs in the event. In the case with the algorithm, on the other hand, there had to be adjustments made to the software parameters to deal with situations as they became clear. Narrative Science

has projected that more than 90% of reporting of this type will be software-generated.

While there are a number of questions that will be raised regarding the impact that these recent adoptions of algorithmically generated and disseminated content have, other questions come from systems and platforms that have a larger scope, as is the case with search engines and commercial platforms, such as Facebook.

2.2.2 Facebook Case Studies

Facebook is a platform with a scope that encompassed over a billion users at the time of this writing. Among the user controls that have been used in relation to data analysis and algorithms is the "Like" button, which signals a vast number of things because it is grossly simplified. With "like" as the only option for rating something on this platform, the user has no way to clearly signify whether they "like the content" in question, "don't like it, but think others should see it," "saw this post," or a number of coded meanings that can result from such a limited option.

It was discovered in 2014 that 62.5% of Facebook users were entirely unaware or unsure that anything specific was governing their feeds (Matias, 2014). Facebook has long been experimenting regularly on the algorithms employed that govern what is now known as the "News Feed" for users. This affects everything about the experience from what connections one might see to what advertisements will be visible to a particular user (Tufekci, 2014).

A controversial experiment was published in 2014 by Adam Kramer and his colleagues, who conducted a study with Facebook data for over 689,000 users, in

which they were able to test if people's reactions to happy posts from friends of Facebook produced "happier" or more "depressed" posts from people on their network. The study determined that happy posts brought forth more happy posts from others on their network.

> "The results show emotional contagion. As [graph] illustrates, for people who had positive content reduced in their News Feed, a larger percentage of words in people's status updates were negative and a smaller percentage were positive. When negativity was reduced, the opposite pattern occurred. These results suggest that the emotions expressed by friends, via online social networks, influence our own moods, constituting, to our knowledge, the first experimental evidence for massive-scale emotional contagion via social networks, and providing support for previously contested claims that emotions spread via contagion through a network" (Kramer, 2014).

A major issue with this was, until that point, it had not previously been common knowledge that anyone *or anything* was curating information and could so masterfully conduct such an experiment without permission or warning. As a result, the scientists experienced a severe backlash including physical threats because of the perceived breach of privacy that Facebook users expressed (Goel, 2014).

Another study focused on user perception regarding the News Feed on their account. The users had a similar initial reaction to learning that there was an algorithm curating content for them.

> "We developed a system, *FeedVis*, to reveal the difference between the algorithmically curated and an unadulterated News Feed to users, and used it to study how users perceive this difference. Participants were most upset when close friends and family were not shown in their feeds. We also found participants often attributed missing stories to their friends' decisions to exclude them rather than to Facebook News Feed algorithm. By the end of the study, however, participants were mostly satisfied with the content on their feeds. Following up with participants two to six months after the study, we found that for most, satisfaction levels remained similar before and after becoming aware of the algorithm's presence, however, algorithmic awareness led to more active engagement with Facebook and bolstered overall feelings of control on the site" (Eslami, 2015).

These case studies offer some factors to consider. One might be compelled to question the lack of transparency at the outset of the deployment of these technologies. If fear is the motivation for "asking for forgiveness rather than asking for permission", in making users unwitting participants in social experiments, what potentially legitimate reasons might a population have to object to such a transition, and under what circumstances do they or do they not have the right to consent to engage in experiments like what Facebook has been running without any regulation or oversight? If emotional contagion is possible through social networks, then what are mental and emotional health guidelines that users need to consider as they engage spaces that have the potential ability to induce positive and negative emotions? If initial findings show promise that informed users make better use of

the platform they engage after understanding some factors of the algorithmic curation, why not make universally known the extent to which machine learning algorithms are part of the environment they occupy and how?

2.3 The Effect of the Algorithm on User Experience

2.3.1 Time and Value on Facebook

Digital mediums can be experienced individually, and at any time. On Netflix or YouTube, for instance, a user can watch any episode of any programming that is present in their database at will, making the events or programming no longer inextricably tied to an "air date" or "air time," which is a significant departure from previous models of media. This departure from a chronological orientation, combined with the algorithms that are coded to curate content based on specific criteria, can alter the perception of time differently. Past, present and future events can all be seen virtually at any time on Facebook, but the algorithm will determine what users can see and from what time.

The Vice President of Product at Facebook shared that "We're in the business of giving our users the most interesting possible experience every time they visit." It was disclosed that there are about 1,500 stories that each user could potentially be shown *every time* they visit FB and view their News Feed. Feasibly, a user can sign into their account simultaneously from two devices and get two entirely different feeds as a result of this preference on the part of the design choices on this platform (McGee, 2013).

One study on a Facebook fan page for a German radio station that had shut down in 1993, DT64, found that the algorithm active during the time of their study rewarded new content over archival content (Kaun & Stiernstedt, 2014). If administrators wanted content to show up on a user's newsfeed, they had to continually post new material rather than simply keep a conversation going in an "older" thread on a post from a week in the past. This imposes an orientation surrounding "liveness" of the interaction that the user is forced to adapt to. In the study, the researchers stated:

> "As Taina Bucher (2012) demonstrated, the EdgeRank algorithm is constructed out of three components: affinity, weight, and time decay. The visibility of content in the newsfeed stream is decided based on the closeness of the relationship between users, the importance of the interaction, and the currency of the post.
>
> Every single post, status update, link, and like in a Facebook feed is visible only for a short period of time: for the user, the experience and feel of Facebook is one of rapid change, new stories are continually appearing, pushing old stories out of sight, downwards in the stream (Keightley, 2012). In the attempt to engage users longer and more often, Facebook rests on the principle of constant change and flow of newness that has been addressed by earlier studies as immediacy and "liveness" of social media" (Bolter et al., 2012; Gerlitz, 2012).

In August of 2013, it was disclosed that the Edge Rank algorithm described above no longer exists in the form as it was, but has been integrated into a ranking algorithm that is far more complex, and is based on machine learning.

> "Lars Backstrom, Engineering Manager for News Feed Ranking at Facebook, estimated that there are as many as '100,000 individual weights in the model that produces News Feed.' The three original EdgeRank elements — Affinity, Weight and Time Decay — are still factors in News Feed ranking, but 'other things are equally important'" (McGee, 2013).

Some of the important factors that the more recent algorithm considered were shared by representatives from Facebook. One is relationship settings, which are volunteered declarations of relationship strength by choosing a setting indicating closeness to another user, rather than those that are extracted from the data. Types of posts a user interacts with and how are a factor, given that some users will be more engaged by photo posts and some may be more engaged by links to articles. Hiding a post or reporting spam also influences what the News Feed algorithm will show any user on the network. Another known factor included in the set of weighted considerations for the algorithm is the type of device used and the strength of Internet connection. A concept called "Story Bumping", which describes the weighted factor called "decay", which can bend the weight rules if an older story is being engaged a certain amount despite how old the content is. One last disclosed factor is called "Last Actor", which describes how the algorithm will factor in with greater weight the last 50 interactions a user had on the network. The weight of the

last 50 interactions of a user will rotate, thus factoring out certain interactions once 50 other interactions have taken place.

III. Data Collection and Why It Matters

3.1 Data Collection Practices

The user-generated content phase of the web consists of users providing their information to the networks they want access to and, in so doing, giving significant amounts of granular data to the provider (Gillespie, 2015). A detail to note is that if the user relinquishes their "ownership" of the data that is amassed on that platform through its user agreement, which is a requirement to "sign" or consent to in some way that is legally binding, then they are no longer able to access that data in any way not required by that agreement.

It is known that every "like," "comment," "share," "message" or other platform affordances have the capacity to inform those looking at the data about specific characteristics of the user. One study showed that researchers clearly could predict with 85% accuracy or more an individual's sexual orientation, race and party affiliation of an individual user just gleaning from "likes" on Facebook. With sufficient crafting of a model, one can reliably determine a user's "sexual orientation, the city, religious and political views, personality traits, intelligence, happiness, use of addictive substances, parental separation, age, and gender" (Kosinski, Stillwell, & Graepel, 2013). Armed with this information, designing appeals to action, marketing, and subtle persuasion is possible. Through the use of computational modeling of data sets as large as what governmental agencies and

commercial platforms like Facebook collect, individual profiles can be generated and catered to based on aspirations, fears, motivations and circumstances that are accurately predictable. One potential application of this modeling is described here:

> "By modeling *individual* psychologies through computational methods applied to big data, a political campaign hoping to garner votes of a conservative candidate can plausibly (and relatively easily) identify voters that were more likely react to fear by voting conservative and vulnerabilities (for example, scares about children for parents; about suffered from accidents; and terrorism, health, or petty crime) while bypassing individuals on whom fear-mongering would not have the desire effect, perhaps even the opposite of the desired one" (Tufekci, 2014).

Thus, "traditional" data gathering, which previously only occurred mainly through mailings, surveys, credit card usage, demographic information and so on, now can be used to cater specific messages to specific users based on this granular data. This greatly increases the likelihood of anticipating the messages that would appeal to individuals across large networks.

3.2 How Much Data is Being Documented?

How much data commercial platforms that have large numbers of users will vary greatly based on the architecture of the platform. Reddit, for instance, uses mostly hyperlinks to information as the means of sharing. As a result, they've reported to be consuming about 2 terabytes of storage total. At an event in August 2012, Facebook shared with attendees "that its system processes 2.5 billion pieces of content and

500+ terabytes of data each day. It's pulling in 2.7 billion 'Like' actions and 300 million photos per day, and it scans roughly 105 terabytes of data each half hour", and that "over 100 petabytes of data are stored in a single Hadoop disk cluster" (Constine, 2012). In order to consider the size of data that is being described, one terabyte is 1,000 gigabytes, and one petabyte is 1,000 terabytes. For comparison, the average full-length feature film in Blue-Ray quality can fit comfortably into a 3 gigabyte file. One estimate made based on calculations from published information is that YouTube is storing around 45 Petabytes per day (Srinivasan, 2014).

IV. Why the Internet is Important to Civic and Political Spheres

4.1 The Internet and Public Spheres

The content generated in and activities facilitated through this global public sphere are as diverse as the users that exist in these spaces. There are benefits and difficulties that come with this. A study from 2014 reports that 67% of the participants had been exposed to hate material online, with 21% of the participants in the study having absorbed and adopted the ideologies of hate groups, the propaganda of which primarily is available on Facebook or YouTube (Oksanen, 2014). A study on mental health found that online support groups are useful for emotional support for the patients as well as assisting healthcare workers in understanding and identifying with patients (Chung, 2014). Events during the Arab Spring were in part coordinated and facilitated effectively through various commercial platforms in addition to the local meetings and operations offline and on the ground to achieve the results that occurred (York, 2012).

4.1.1 Solidarity and Activism

One benefit from the scope of the public sphere having expanded beyond county, state and national boundaries into this global community is that users with access can publish content that is unique to their locality, making such content available internationally for those who may not have the ability to travel to that

physical location. This benefit also extends to those who wish to be aware of and support in some way groups who experience oppression or who have decided to take action in some way counter to a social institution that they either feel is unjust or not something they can support or identify with. A few international examples embody such solidarity or activist movements that made use of commercial platforms to share a message with each other, and in so doing to a global public sphere, a global audience.

Invisible Children's KONY 2012 video, which was produced to make an appeal for attention and aid for child soldiers being abducted in Uganda, received attention and achieved numbers that many had not considered possible. Prior to that video, the most popular YouTube videos for films were at around 200,000 to 300,000 hits, and KONY 2012 reached 43,000,000 hits within 72 hours (Chalk, 2012). The proliferation of cell phone cameras and access to platforms like Twitter to publish content enabled the minority population of Ferguson, Missouri, to expose a systemic oppression and corruption on the part of the police and other governmental institutions in Ferguson, the results of which were later released in a document generated by the State Department of the United States.

Many more examples exist of civic engagement on various platforms in different ways and for different reasons depending on the architecture and user affordances.

4.1.2 Identity and Anonymity

It is worth touching upon the topic of identity and anonymity in these online spaces, for not every user of social media desires to have their activities on the web

present a direct duplicate of their offline persona, but the policies of various online spaces may or may not allow this to be decided by the user. There are various reasons for employing anonymity; such as working as a journalist or activist, where personal well-being and the well-being of family can be put in danger as a result of such exposure; or an assumed identity protects a person from reprisals from their local community for having different interests, ideas or beliefs.

The topic of anonymity and identity has been a point of challenge and conflict for centuries. Classic literary works, such as Robinson Crusoe, Pride and Prejudice and Frankenstein were published anonymously for a variety of reasons. Charles Lutwidge Dodgeson was so shy that he wrote Alice in Wonderland under the pen name of "Lewis Carroll." Some legitimate reasons could be spiritual choices, where a person having their name be praised, instead of God is undesirable. There are women who chose to publish under a male name so that the work would be read without gender bias as well as many authors that felt that the frankness and honesty of their writing would cause trouble with the governmental forces of the time (Stryker, 2012).

Facebook is one public space that has a relatively strict identity policy, which has been the subject of controversy. Mark Zuckerburg, CEO of Facebook, in an interview said that, "You have one identity," and that essentially "The days of you having a different image for your work friends or co-workers and for the other people you know are probably coming to an end pretty quickly." He continues, "Having two identities for yourself is an example of a lack of integrity," which has met with resistance from users who either value their anonymity or their identity

has shifted to a name other than their government-issued identification card. An issue arose in October of 2014, when members of the LGBT community attempted to use their most common and preferred identity for their account and were denied. Facebook has since "apologized to the drag queens and the broader queer community supporting them. Their policy of requiring only 'real names' in users' profiles will now allow whatever "authentic name they use in real life" (Gillespie & Lingel, 2014).

A phenomenon directly related to issues of identity and anonymity online is the concept of "trolling", which has been said to have been invented and popularized on a space on the web called "4Chan," in a particular subdirectory where a variety of content is shared and people gather (Stryker, 2012). The definition of the act of trolling varies somewhat, but as this definition states:

> "Trolling is a game about identity deception, albeit one that is played without the consent of most of the players. The troll attempts to pass as a legitimate participant, sharing the group's common interests and concerns [...] A troll can disrupt the discussion on a newsgroup, disseminate bad advice, and damage the feeling of trust in the newsgroup community" (Karppi, 2013).

The concept and practice of trolling moved beyond less public spaces to commercial platforms like Facebook, Twitter, YouTube, Comments sections for news articles and virtually any place where the "game" mentioned above might be played. In fact, there have been repeated occurrences of memes that suggest changing one's profile name to "No one" and go around clicking on the Facebook "like" button on other

user statuses in order to send the message that "No one likes" their status. While the members of the drag community argued that their adopted identities were in fact how they identified themselves in real life, preventing trolling behavior falls directly under the sentiment behind this Facebook policy:

> "We believe that using your real name, connecting to your real friends, and sharing your genuine interests online create more engaging and meaningful experiences. Representing yourself with your authentic identity online encourages you to behave with the same norms that foster trust and respect in your daily life offline. Authentic identity is core to the Facebook experience, and we believe that it is central to the future of the web. Our terms of service require you to use your real name and we encourage you to be your true self online, enabling us and Platform developers to provide you with more personalized experiences" (Karppi, 2013).

The conversation about identity and anonymity is central to how the public sphere is observed from within and without. If all accounts, pictures, travel dates, Instagram tags, credit card purchases, GPS searches and status updates match up with the account name, there's a digital coherence that can be attained to gather a clear snapshot of each individual user.

4.2 The Internet and Political Spheres

According to the *World Factbook* from the United States Central Intelligence Agency, over 3 billion people are online worldwide as of 2013. Public social spaces are widely used from this online population. As of March of 2015, Facebook reports

1.39 billion users, Twitter reports 288 million monthly active users, and YouTube reports over 1 billion users with hundreds of millions of hours watched daily, which is just to name three of the more visible commercial platforms where people gather in this online public sphere.

Public spheres where citizens gather, wherever that may be, is of interest to governmental institutions, be they just to remain in touch with the sentiment of the people they govern, or to engage in surveillance to collect data. Governments have been known to shut down certain sites, such as DailyMotion in Tunisia in 2007, but there is collateral damage that creates a dilemma. Ethan Zuckerman coined "The Cute Cat Theory of Digital Activism" in 2007 to describe a public platform that is difficult to censor without also censoring non-political content, which makes the use of non-political platforms for activism and political activity less likely to be censored in such public spaces, which would bring public attention to such censorship (Zuckerman, 2014). Though his description was not specifically limited to online platforms, Clay Shirky's (2011) description of this phenomenon is apt:

> "…the conservative dilemma," so named because it applies not only to autocrats but also to democratic governments and to religious and business leaders. The dilemma is created by new media that increase public access to speech or assembly; with the spread of such media, whether photocopiers or Web browsers, a state accustomed to having a monopoly on public speech finds itself called to account for anomalies between its view of events and the public's. The two responses to the conservative dilemma are censorship and propaganda. But neither of these is as effective a source of control as the

> enforced silence of the citizens. The state will censor critics or produce propaganda as it needs to, but both of those actions have higher costs than simply not having any critics to silence or reply to in the first place. But if a government were to shut down Internet access or ban cell phones, it would risk radicalizing otherwise pro-regime citizens or harming the economy" (p. 6).

4.2.1 Voting Behavior

One study concluded that specific age groups form their opinions and preferences on Facebook and Twitter, and political candidates should develop a capacity to effectively use multiple platforms to "triangulate opinions on the way to the voting booth" as a political strategy. (Peterson, 2012) In a 2014 study, The Pew Research Center stated:

> "The proportion of Americans who use their cell phones to track political news or campaign coverage has doubled compared with the most recent midterm election: 28% of registered voters have used their cell phone in this way during the 2014 campaign, up from 13% in 2010. Further, the number of Americans who follow candidates or other political figures on social media has also risen sharply: 16% of registered voters now do this, up from 6% in 2010."

What becomes clear from these findings is that while entities like political candidates and governmental institutions are using their access to the public sphere to study and potentially influence the public using these platforms, they are also *required* to engage in these ways in order to maintain a competitive edge.

Candidates that don't post information to potential voters with regularity and in the spaces that are being occupied by them on these commercial platforms, they are at risk of losing elections as a result. Governments have to spend significant resources to continue to be current with the spheres where the most genuine conversations are being had in order to maintain an understanding of citizens.

The political sphere is also subjected to public scrutiny and being quickly put in the spotlight. Legislations and activities that might have been limited in view from the traditional media outlets of newspaper and network news before the advent of the Web now can be broadcast to a global audience in a short time. Online activism has an influence on this sphere when it comes to public information, networks and opinion. In May 2011, the Protect IP Act (PIPA), was proposed by a Vermont senator with 11 cosponsors to the US Senate. The bill was also officially known as "the Preventing Real Online Threats to Economic Creativity and Theft of Intellectual Property Act" which as the title suggests, would enable various forms of enforcement and accountability online to things considered to be proprietary. Through various channels, online and offline, proponents of free speech, uncensored Internet, generators and consumers of web content and many other groups began a campaign to publicly broadcast the issues with such a legislation. Wikipedia, for instance, had a 24-hour blackout on their website, and 8 million US visitors contacted their Congressional representatives. Google had a petition to Congress against PIPA and SOPA, which was signed by 4.5 million users. Within 8 months, the PIPA Act went from proposal to being indefinitely postponed (O'Leary, 2012).

The 2012 Presidential Election demonstrated a clear shift in the political sphere in online spaces. Every statement has the potential to amplify and become the viral quote of the day on platforms that have billions of interconnected users. The following analysis describes pressure on political figures as a result:

> "Now with online campaigns, perfection is not only needed when appearing on television, but at all times. The modern campaign requires constant composure on the part of the candidate because any slip up will be splashed across the headlines of a website instantly.
>
> Online news has taken over politics and has allowed every voter the chance to become a political analyst. Mitt Romney discovered the instant backlash of online news after he made a comment about the 47% of Americans who don't pay income taxes, as well as his "binders full of women" remark. The immediate response came in tweets, comments, and Facebook posts, but the delayed reaction featured bloggers posting edited images mocking the Romney lines. An entire blog on Tumblr was devoted to featuring joking images of everyone's interpretations of binders full of women" (Dalton-Hoffman, 2013).

Both Mitt Romney and Barack Obama spent close to $100,000 on just their data based online advertisements. Both candidates hired third party companies to track data on their campaign websites. Ann Romney shared with the world on Pinterest her favorite kind of bread, a White House security guard playing Jenga and an assortment of snapshots from the candidates' daily lives were displayed publicly as ways to relate to the public and seem accessible.

V. How Platform Algorithms Are Important to Civic and Political Spheres

5.1 Algorithmic Curation

Algorithmically curating human preferences has been noted to be a problematic, albeit convenient, solution to engaging a diverse audience with a common experience. The frequency with which a Facebook Newsfeed displays a baby picture, something from yesterday, something from a close friend, a distant relative, something from another country; all of these decisions are computer-generated. The data is inputted by humans and the programming is generally done by humans, but the calculations made by the algorithm, the output, can have some damaging unintended consequences. Alex Leavitt (2014) describes the emotional pain of a friend who was presented with a picture of his daughter who had died that year, prominently displayed as the cover photo for his Facebook "Year in Review" video. The video came with a tag-line, "It's been a great year! Thanks for being part of it." The author noted his own emotional struggles with depression and how Facebook's algorithm delivered a painful reminder of the seeming gaps in his life. He went on to analyze the data that he was able to reverse engineer from Facebook using his coding abilities, and tried to understand how the algorithm chose the events that it did. One conclusion drawn centered around "Attention Economics", which in choosing from a collection of life events, those that attract the most

attention are considered to be "more important" or even "preferable" to less sensational moments. He describes some events that had been chosen in the past:

> "Not surprisingly, the 20 most important events of my year were vacations and photos with celebrities—the events that received the most likes and comments from my Facebook friends and subscribers. I did notice that my events were present-heavy—four events from the past week made my Year in Review" (Leavitt, 2014).

Another author who is an active Twitter user critiqued a proposed algorithmically curated Twitter Feed, which would replace the chronological timeline that hasn't changed since Twitter launched, and would be a change to the core of the platform experience. She went on to explain that, admittedly, this change would likely reward her as a long-standing user with a wide network as far as how visible she might be on an algorithmically curated feed. Her objections to the change, however, were because it would remove a human-interpretation element mixed with the setting of a wide network that could extrapolate unexpected things from "networked intelligence," and deliver little more than any other popular media outlet. To highlight the contrast in approach, she describes what could be lost from a curated feed:

> "But the bigger loss will be the networked intelligence that prizes emergence over engagement and interaction above the retweetable— which gets very boring very quickly. I know Twitter thinks it may increase engagement, but it will decrease engagement among some of its most creative segments" (Tufekci, 2014).

5.2 User Awareness of Local, National and International Events and Phenomena

Algorithmic filtering, as was described above, will output either over-emphasis on one type of content or suppress another kind, depending on the categorization of interests and priorities of the user and designer. The weight of that categorization in opaque algorithms is difficult to know with certainty. In the case of Ferguson, Missouri, after an unarmed young man named Michael Brown was killed by a police officer in August of 2014. There were protests for many days. Twitter activity picked up that these protests were occurring, and video footage and on-the-ground commentary began circulating. At the same time, many users were looking for mention of this event on Facebook and found none. A likely scenario is that Facebook's algorithm hadn't yet made associations with mentions of "Ferguson" or "Michael Brown" as newsworthy or worthy of attention (Tufekci, 2014).

This event, and other local, national and international events that occur may not register as "important", "newsworthy" or "exciting" by the standards of certain portions of the population. One of the most powerful things about a global community meeting in spaces together is the "networked intelligence", or the collective strengths that individuals within a network experience when pooling together their collective knowledge, that comes from the collective power of strengths and a scaffolding to compensate for any shortcomings that any one group may suffer from. If there are insurmountable filters between all actors that intend to understand each other; if there are prioritization decisions deciding for users who is

"recommended" to engage; there may be consequences of adopting systems that serve us in certain capacities, but bring unknown challenges with them.

5.3 Marketing and Advertising

Wherever the public sphere conglomerates, there is the interest of those who have goods or services to trade or sell. The commodification of public attention and information in online spaces, is of note in light of how significantly the commercial market has been impacted by Web 2.0. Brick and mortar businesses that once were regarded as household names like Blockbuster perished in the fast-paced landscape and shifting paradigm away from physically-based spaces being the dominant sources for commerce (Michel, 2008). The most commercially successful online platforms to date have made access to users of the platform free of charge in exchange for control of data, privacy and even the virtual identity cultivated on the platform.

This has often come in the form of advertisements on user profiles and News Feeds on various platforms. Among the platforms that have made use of connecting advertising to users, Facebook has been very visible. Representatives from Facebook made clear some things about the advertising algorithms that operate on the platform.

> "The News Feed algorithm is completely separate from the algorithm that decides what ads to show, when to show ads, and where to show them. But how a user interacts with Facebook ads *can* influence what shows in the News Feed.

> 'Nothing is off the table when we're looking at what we should show users,' Lars Backstrom, Engineering Manager for News Feed Ranking at Facebook says. 'It can be clicking on ads or looking at other timelines. It doesn't have to be just what the user interacts with in the News Feed.'" (McGee, 2013, p.)

An announcement in June of 2014 brought the news that Facebook will expand the parameters of how it will share user to include commercial spaces outside of Facebook.

> "If a Facebook user researches a new television on an external website or inside of a mobile app, their profile might now indicate an interest in televisions and in electronics, making it easier for advertisers pitching electronic devices to reach that user on Facebook" (Oreskovic, 2014, p.).

Access to user profiles in ways that they would be most receptive to the messages is the goal for advertisers, and a platform with such large amounts of individual user data is in the position to safeguard the data and privacy of users if they so choose. It would seem that there are ethical questions to be asked regarding the selling and releasing of individual user data and exposing that to third parties, but at present there is no legal responsibility on the part of Facebook to the users on the network.

VI. Influence

6.1 How the Structure of the Platform Influences the User

How the platform is constructed can concretely impact the user's perception of outside events.

> "…the methods of transmitting information are decided by engineers who do not exist in a social, moral or psychological vacuum, and that the decisions of these engineers can have enormous consequences (both deliberate and unforeseen) on society, consequences which can only be affected by law when their creation, deployment and utilisation are understood and where the other forces of economics and psychology/sociology are also considered" (Adams, 2014).

The "Emotional Contagion" study that Facebook internally conducted and published in 2013, made clear that user posts are impacted and influenced to some degree by the nature of their interactions with their network (Kramer, 2014). In what ways might users be affected by the architecture of the platform? What choices went into the decisions that users can and cannot make about their privacy?

In a study of 61 Million Facebook (FB) users in 2010 conducted by Robert Bond and colleagues, it was discovered that one social message had the power to increase voter turnout of an election in the 2010 congressional elections. This number is notable when one considers that the Bush/Gore 2000 Presidential

Election came down to 537 votes, which amounts to less than 0.01 % of votes in Florida.

> "The social message group (n = 60,055,176) was shown a statement at the top of their News Feed. This message encouraged the user to vote, provided a link to find local polling places, showed a clickable button reading I Voted, showed a counter indicating how many other Facebook users had previously reported voting, and displayed up to six small randomly selected 'profile pictures' of the user's Facebook friends who had already clicked the I Voted button. The informational message group (n =611,044) was shown the message, poll information, counter and button, but they were not shown any faces of friends. The control group (n =613,096) did not receive any message at the top of their News Feed" (Bond, Fariss, Jones, Kramer, Marlow, Settle & Fowler, 2012).

The researchers altered only one element with a few variations on Facebook that could be visible to achieve that result. In changing one element in a social network with that scope of users and tracking three user affordances - clicking the "I voted" button, clicking the polling place link and voting in the election - the researchers were able to effectively increase voter participation by a total of 340,000 votes. The growth in turnout between 2006 and 2010 of 0.60% may well have been caused by one message disseminated on Facebook (Bond et. al, 2012).

While this study was not a comparative analysis of FB against other platforms, a few things might be deduced from this 2010 election study. One is that the platform affordances of Facebook differ from other platforms in respect to the

influence that one message can have throughout its network of users. In their study it was stated: “Our validation study shows that close friends exerted about four times more influence on the total number of validated voters mobilized than the message itself” (p. 5). It was also noted that, “messages including cues from an individual’s social network are more effective than information-only appeals,” and that “close friendships accounted for all of the significant contagion of these behaviors, in spite of the fact that they make up only 7% of all friendships on Facebook.” (p. 4) On a different type of social network, such as Reddit, the same influence that might come from the trust a user may have in a network that contains users from their off-line life as well was not as readily present (Bond, 2012).

Another thing to be deduced is that Facebook as a public sphere can be a particularly valuable resource for learning what diverse groups find important, understanding the codes in which they communicate with each other, and make appeals to such groups with a precision that was not possible before the advent of Web 2.0.

6.2 Algorithms and Their Influence

There is much literature on the potential uses of algorithms for influence (Romero, 2011), for identification (Becker, Namaan & Gravano, 2010) and for making the user experience more convenient (Soman & Murugappan, 2014), but until the “Emotional Contagion” study was published, the public consciousness was not focused on how it might be affecting their lives. Facebook alone has very detailed data on individuals as well as how they are connected to large numbers of people also in the Facebook network. Twitter and some other social networks

operate in similar ways, with Twitter being the most widely researched because of their open data and simple set of user controls (Tufekci, 2014).

6.2.1 Influence of Values

The more any user and an algorithm interact, the likelihood that the algorithm will move toward a feedback-loop of homogeneity, described by Eli Pariser as “filter bubbles”, rises. A filter bubble occurs when an algorithm personalizes the user experience so effectively that the user eventually occupies a space, and is exposed *exclusively* to ideas and other users that are consistent with their own values, perspectives and beliefs (Pariser, 2011). In such filter bubbles, users can develop the mistaken impression that “everyone knows this” or that a disproportionate segment of the population agrees with a certain set of values or ideas.

Gillespie describes algorithmically generated *calculated publics,* in which the algorithm according to design makes associations based on available data, and assumes that a user belongs within a certain grouping.

> “Algorithms not only structure our interactions with others as members of networked publics; algorithms also traffic in *calculated publics* that they themselves produce. When Amazon recommends a book that "customers like you" bought, it is invoking and claiming to know a public with which we are invited to feel an affinity -- though the population on which it bases these recommendations is not transparent, and is certainly not coterminous with its entire customer base. When Facebook offers as a privacy setting that information can be seen by "friends, and friends of

> friends," it transforms a discrete set of users into an audience -- it is a group that did not exist until that moment, and only Facebook knows its precise membership. These algorithmically generated groups may overlap with, be an inexact approximation of, or have nothing whatsoever to do with the publics that the user sought out" (Gillespie, 2014).

In the sphere of algorithmic influence, users are required to consider whether or not they belong in the calculated public that was suggested in some cases, and chosen for them in others. When Netflix suggests to a user that "Because You Watched…You May Want to Watch" another selection of their choosing, the question remains whether that is true.

Christian Sandvig identified some concrete challenges associated with the growing permeation of algorithmic curating. He describes a concept that he calls "Corrupt Personalization", in which "your attention is drawn to interests that are not your own." A few points were shared for consideration:

> "When I express my opinion about something to my friends and family, *I do not want that opinion re-sold* without my knowledge or consent. When I explicitly endorse something, *I don't want that endorsement applied to other things* that I did not endorse. If I want to read a list of personalized status updates about my friends and family, *I do not want my friends and family sorted by how often they mention advertisers.* If a list of things is chosen for me, I want the results organized by some measure of goodness for me, *not by how much money* someone has paid. I want *paid content* to be clearly

> identified. I do not want my information technology to sort my life into commercial and non-commercial content and *systematically de-emphasize the noncommercial things that I do,* or turn these things toward commercial purposes" (Sandvig, 2014).

Along with the above assertions, Sandvig described dangers associated with a system defaulting to an economic orientation, and thus operating upon approaches and values that present content through a commercial filter. From this perspective, "lucrative" or "profitable" are equitable with "best," rather than the wide variety of experiences and qualities of the human experience that have no quantifiable monetary value. Long-term exposure to this filter also has the possibility to be normalized and suggestive to the point where users are effectively "taught what to want" without a developed capacity to question the truth of the suggestion. All of these effects are potentially the outpourings of "the economic organization of the system" that we are considering.

> "With algorithmic culture, computers and algorithms are allowing a new level of real-time personalization and content selection on an individual basis that just wasn't possible before. But rather than use these tools to serve our authentic interests, we have built a system that often *serves a commercial interest that is often at odds with our interests* — that's *corrupt personalization.*
>
> If I use the dominant forms of communication online today (Facebook, Google, Twitter, YouTube, etc.) I can expect content customized for others to use my name and my words without my consent, in ways I wouldn't approve

of. *Content "personalized" for me includes material I don't want, and obscures material that I do want.* And it does so in a way that I may not be aware of" (Sandvig, 2014).

As was discovered in the "61 Million Facebook user study" of the 2010 congressional election, there is clear and documented evidence that 340, 000 voters were influenced to participate in the voting process directly from one element on an online platform (Bond, 2012). It cannot be assumed that the principles and values of the system are incorruptible, free from bias nor inherently moral or ethical. To what might we owe the confidence that any platform algorithm will not be adjusted to make one candidate more visible than another? With such algorithms that have a reach of billions, yet are in a "black box", how might such an adjustment be discovered or proven?

6.2.2 Discriminatory Influence

In the offline world in the United States, there are laws within a legal system that offers ways for social contracts and agreements to be upheld for the protection of citizens of a locality, state or nation. Among these laws are those that are intended to protect citizens from discriminatory practices on the part of individuals, communities or institutions. In physical spaces, there are precedents for many cases that would be argued by the defense and the prosecution. Online platforms occupy a different landscape that has proven difficult to adjudicate law in the same way (Kerr, 2003).

Racial discrimination, which has a long and complicated history in the US, is an issue that has a number of laws regarding hiring practices, exclusion and access

to resources for people of various cultures and races that are not of European descent. Online spaces presently have limitations regarding what one can pinpoint to be the "source" of discrimination, particularly in the case of algorithmic output. If an algorithm is in a "black box", then the data and weights assigned to any possible associations made from the data are difficult to identify, let alone prove. For example, one researcher found that Google algorithms were returning queries with ads that suggest discrimination.

> "A greater percentage of ads having "arrest" in ad text appeared for black identifying first names than for white identifying first names in searches on Reuters.com, on Google.com, and in subsets of the sample. Results of Chi-Square tests on these patterns were statistically significant. On Reuters.com, a host of Google AdSense ads, a black-identifying name was 25% more likely to get an ad suggestive of an arrest record, $X^2(1)=14.32$, $p < 0.001$; there is less than a 0.1% probability that these data can be explained by chance" (Sweeny, 2013).

Systems are developing with algorithms that can make hiring decisions, decisions whether to offer loans and decisions to approve or deny health care among many other services. If the data being calculated regarding the decisions that these systems are making can not be viewed or analyzed, the binary response, "approved" or "denied" cannot be subject to inquiry regarding discriminatory practices (Rosenblat, 2014).

> "Hiring algorithms are designed using historical data to create predictions about which qualities correlate to a strong job performance. If

> workplace discrimination has historically elevated the performance of one group over others, then algorithms derived from such historical data will tend to reinforce that historical bias. This is likely even if the algorithm's designer does not intend to discriminate based on these categories. Even without such historical bias, the accuracy of an algorithm's outputs will still be higher for the dominant statistical group, and lower for the statistical minority, because more information is available about job performance for the larger group. Algorithms that are trained on data sets that include both inaccurate and accurate information output biased or meaningless information; worse, such dirty data can dilute other valuable signals and send inaccuracies rippling through entire systems. Because of limited transparency, unclean data, and the complexity of most algorithms used to do this kind of analysis, it is often difficult to discern the specific reasons for which a job candidate receives a negative score" (Rosenblat, 2014).

The implications of this sphere of influence is significant, and research in this area is illuminating the need for inquiries like the ones described above to move forward as swiftly as the technological advances that are being applied in various sectors of political and civic life.

CONCLUSION

The results of this analysis bring to light the limitations of the algorithms that platforms implement and implications of the lack of diversity of data included in them. It makes clearer the lack of user affordances to control the information they wish to see or not, and how that challenge is compounded in a network of users regarding what they see and can effectively share within networks. Ultimately, it sheds some light into how we may need to think of these tools in the future: not so much as commercial engines for economic growth, but more so engines to develop agency within users to become empowered participants in social connectivity, growth and global collaboration.

APPENDIX

The majority of the literature found in this review came from the Google Scholar Database and some from the ACM Digital Library using search keywords:

- Social media
- Social media society
- authority social media
- Algorithm social media
- Solidarity social media
- Twitter Facebook algorithm
- Marketing social media
- Online advertising
- Algorithm curation
- YouTube User Affordances
- Facebook User Affordances
- Twitter User Affordances
- Reddit user
- Tumblr user
- Reddit affordance
- Tumblr affordance
- Facebook vote 2012
- Twitter vote 2012
- Tumblr vote 2012
- Reddit vote 2012

Method

Analysis

The research question was about the interaction between platform affordances through which users can signal their preferences and interests, platform algorithms through which content gets amplified or dampened, and how these policies and design choices impact political and civic arenas. The approach selected for this paper was to analyze existing research on platform algorithms and platform affordances and illuminate the ways in which they can obscure or shed light upon information on behalf of individual and collections of users within its domain without their full awareness of it.

The researcher collected a sampling of available journals on platform affordances and algorithms to analyze their design. Then available information was gathered on the affordances and algorithms for Facebook, Twitter, YouTube, Tumblr and Reddit to explore the effects of each platform's opacity or transparency and its effect on users. Because these platforms represent varying scope of membership, user affordances and transparency to the user, this analysis should offer a useful cross-study of how they impact political and civic arenas.

Bibliography

Adams, A. A. (2014). *Facebook code: Social network sites platform affordances and privacy*.

Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D., Marlow, C., Settle, J. E., & Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. *Nature*, *489*(7415), 295-298.

Burgess, J. (2014). From 'Broadcast yourself' to 'Follow your interests': Making over social media. *International Journal of Cultural Studies*, 1367877913513684.

Chalk, S. (2012). KoNy 2012: SuCCESS oR fAILuRE?. *London: International Broadcasting Trust*.

Choy, M., Cheong, M., Laik, M. N., & Shung, K. P. (2012). US presidential election 2012 prediction using census corrected twitter model. *arXiv preprint arXiv:1211.0938*.

Chung, J. E. (2014). Social networking in online support groups for health: how online social networking benefits patients. *Journal of health communication*, *19*(6), 639-659.

Clerwall, C. (2014). Enter the Robot Journalist. *Journalism*, *8*(5).

Constine, J. (2012, August 22). How big is Facebook's data? 2.5 billion pieces of content and 500+ terabytes ingested every day. *TechCrunch.* Retrieved from http://techcrunch.com/2012/08/22/how-big-is-facebooks-data-2-5-billion-pieces-of-content-and-500-terabytes-ingested-every-day/

Crawford, S. (2014). *The cliff and the slope.* Retrieved from https://medium.com/backchannel/jammed-e474fc4925e4

Dalton-Hoffman, M. (2013). *The effect of social media in the 2012 Presidential election.* Retrieved from http://www.digitalrepository.trincoll.edu

Darmon, D., Sylvester, J., Girvan, M., & Rand, W. (2013, September). Predictability of user behavior in social media: Bottom-up v. top-down modeling. In *Social Computing (SocialCom), 2013 International Conference on* (pp. 102-107). IEEE.

Dewey, C. (2014, October 29). *Almost as many people use Facebook as live in the entire country of China*. *Washington Post.* Retrieved from http://www.washingtonpost.com/news/the-intersect/wp/2014/10/29/almost-as-many-people-use-facebook-as-live-in-the-entire-country-of-china/

Dey, R., Jelveh, Z., & Ross, K. (2012, March). Facebook users have become much more private: A large-scale study. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2012 IEEE International Conference on* (pp. 346-352). IEEE.

Diakopoulos, N. (2014). Algorithmic accountability: Journalistic investigation of computational power structures. *Digital Journalism*, (ahead-of-print), 1–18.

Diakopoulos, N., Cass, S., & Romero, J. (2014, October). Data-driven rankings: The design and development of the IEEE top programming languages news app. In *Proc. Symposium on Computation+ Journalism*.

Duggan, M., Ellison, N.B., Lampe, C., Lenhart, A., & Madden, M. (2015, January 9). *Social Media Update 2014*. Pew Research Center, January 2015. Retrieved from http://www.pewinternet.org/2015/01/09/social-media-update-2014/

Duhigg, C. (2012, February 16). How companies learn your secrets. *New York Times.* Retrieved from http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?pagewanted=1&_r=1&hp

Dynel, M. (2014). Participation framework underlying YouTube interaction. *Journal of Pragmatics*, *73*, 37-52.

Edburg, J. (2013, August 26). *Reddit: Lessons learned from mistakes made scaling to 1 billion pageviews a month.* Retrieved from http://highscalability.com/blog/2013/8/26/reddit-lessons-learned-from-mistakes-made-scaling-to-1-billi.html

Ellison, N. B., & Vitak, J. (2015). Social network site affordances and their relationship to social capital processes. *The Handbook of the Psychology of Communication Technology*, *33*, 205.

Eltantawy, N., & Wiest, J. B. (2011). The Arab Spring| Social Media in the Egyptian Revolution: Reconsidering resource mobilization theory. *International Journal of Communication*, *5*, 18.

Eslami, M., Rickman, A., Vaccaro, K., Aleyasen, A., Vuong, A., Karahalios, K., ... & Sandvig, C. (2015). "I always assumed that I wasn't really that close to [her]": Reasoning about invisible algorithms in the news feed. CHI.

Flew, T., Spurgeon, C., Daniel, A., & Swift, A. (2012). The promise of computational journalism. *Journalism Practice*, *6*(2), 157-171.

Fox, J., & Moreland, J. J. (2015). The dark side of social networking sites: An exploration of the relational and psychological stressors associated with Facebook use and affordances. *Computers in Human Behavior*, *45*, 168-176.

Gilbert, E. (2013, February). Widespread under provision on Reddit. In *Proceedings of the 2013 conference on Computer supported cooperative work* (pp. 803-808). ACM.

Gillespie, T. (2014). 9 The relevance of algorithms. *Media Technologies: Essays on Communication, Materiality, and Society*, 167.

Gillespie, T., Lingel, I. (2014, October 2). *One name to rule them all: Facebook's identity problem.* Retrieved from http://www.theatlantic.com/technology/archive/2014/10/one-name-to-rule-them-all-facebook-still-insists-on-a-single-identity/381039/

Goel, V. (2014). 'Facebook tinkers with users emotions in news feed experiment, stirring outcry. *New York Times, June*, *29*.

Gross, R., & Acquisti, A. (2005, November). Information revelation and privacy in online social networks. In *Proceedings of the 2005 ACM workshop on Privacy in the electronic society* (pp. 71-80). ACM.

Hampton, K. N., Goulet, L. S., Marlow, C., & Rainie, L. (2012). Why most Facebook users get more than they give. *Pew Internet & American Life Project*, *3*.

Hight, C. (2014). Shoot, edit, share: Cultural software and user-generated documentary practice. *New Documentary Ecologies: Emerging Platforms, Practices and Discourses*, 219.

Hillman, S., Procyk, J., & Neustaedter, C. (2014, June). 'alksjdf; Lksfd': tumblr and the fandom user experience. In *Proceedings of the 2014 conference on Designing interactive systems* (pp. 775-784). ACM.

Houghton, D., Joinson, A., Caldwell, N., & Marder, B. (2013). *Tagger's delight? Disclosure and liking in Facebook: the effects of sharing photographs amongst multiple known social circles.*

Jeong, Y., & Coyle, E. (2014). What are you worrying about on Facebook and Twitter? An empirical investigation of young social network site users' privacy perceptions and behaviors. *Journal of Interactive Advertising*, *14*(2), 51-59.

Karppi, T. (2013). FCJ-166 'Change name to No One. Like people's status' Facebook Trolling and Managing Online Personas. *The Fibreculture Journal*, (22 2013: Trolls and The Negative Space of the Internet).

Kaun, A., & Stiernstedt, F. (2014). Facebook time: Technological and institutional affordances for media memories. *New Media & Society*, *16*(7), 1154-1168.

Keightley, E. (Ed.). (2012). *Time, Media and Modernity*. Palgrave Macmillan.

Kerr, O. S. (2003). The problem of perspective in Internet law. *Georgetown Law Journal*, *91*.

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, *110*(15), 5802-5805.

Kramer, A. D., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, *111*(24), 8788–8790.

Leavitt, A. (2014, December 31). *Reviewing my facebook year in review: Algorithmic curation, emotional traces, and values in design*. Retrieved from https://medium.com/i-data/reviewing-my-year-in-review-algorithmic-curation-trace-data-and-values-in-design-938ae2e3489b

Lessig, L. (1999). *Code and other laws of cyberspace*. Basic books.

Levy, S. (2012, April 24). *Can an Algorithm Write a Better News Story Than a Human Reporter?* Wired. Retrieved from http://www.wired.com/2012/04/can-an-algorithm-write-a-better-news-story-than-a-human-reporter/

Maximilien, E. M., Grandison, T., Sun, T., Richardson, D., Guo, S., & Liu, K. (2009, May). Privacy-as-a-service: Models, algorithms, and results on the facebook platform. In *Proceedings of Web* (Vol. 2).

McGee, M. (2013, August 16). *EdgeRank is dead: Facebook's news feed algorithm now has close to 100k weight factors*. Retrieved from http://marketingland.com/edgerank-is-dead-facebooks-news-feed-algorithm-now-has-close-to-100k-weight-factors-55908

Michel, S., Brown, S. W., & Gallan, A. S. (2008). An expanded and strategic view of discontinuous innovations: deploying a service-dominant logic. *Journal of the Academy of Marketing Science*, *36*(1), 54-66.

Miller, R. (2015, January 29). *AP's "robot journalists" are writing their own stories now.* Retrieved from http://www.theverge.com/2015/1/29/7939067/ap-journalism-automation-robots-financial-reporting

Morrison, A. (2014). Facebook and coaxed affordances. *Identity Technologies: Constructing the Self Online*, 112.

Murthy, D. (2013). Twitter. *Digital Media and Society Series*.

Obar, J. A. (2014). Canadian advocacy 2.0: An analysis of social media adoption and perceived affordances by advocacy groups looking to advance activism in Canada. *Canadian Journal of Communication*, *39*(2), 211-233.

Oksanen, A., Hawdon, J., Holkeri, E., Näsi, M., & Räsänen, P. (2014). Exposure to online hate among young social media users. *Soul of Society: A Focus on the Lives of Children & Youth (Sociological Studies of Children & Youth, volume 18)*, 253-273.

O'Leary, D. E. (2012). Computer-based political action: The battle and Internet blackout over PIPA. *IEEE Computer*, *45*(7), 64-72.

Olson, J. S., & Kellogg, W. A. (2014). *Ways of knowing in HCI*. Springer, New York, NY: URL: 〈 http://link.springer.com/10.1007/978-1-4939-0378-8.

Oreskovic, A. (2014, June 12). *Facebook will sell data on its users' online activities outside of Facebook.* Retrieved from https://bangordailynews.com/2014/06/12/business/facebook-to-begin-selling-data-on-its-users-online-activities-outside-of-facebook/

Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. UK: Penguin.

Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press.

Pennington, N., Winfrey, K. L., Warner, B. R., & Kearney, M. W. (2015). Liking Obama and Romney (on Facebook): An experimental evaluation of political engagement and efficacy during the 2012 general election. *Computers in Human Behavior*, *44*, 279-283.

Pepitone, J., & January, C. (2012). SOPA and PIPA postponed indefinitely after protests. *CNNMoney.com Retrieved from: http://money.cnn.com/2012/01/20/technology/SOPA_PIPA_postponed/index.html*

Pew Research Center (PRC), November, 2014, "Cell Phones, Social Media, and Campaign 2014" Available at: http://www.pewinternet.org/2014/11/03/cell-phones-social-media-and-campaign-2014/

Podolny, S. (2015, March 7). *If an Algorithm Wrote This, How Would You Even Know?* Retrieved from http://mobile.nytimes.com/2015/03/08/opinion/sunday/if-an-algorithm-wrote-this-how-would-you-even-know.html?_r=1&referrer=

Porter, A. J., & Hellsten, I. (2014). Investigating participatory dynamics through social media using a multideterminant "frame" approach: The case of climategate on YouTube. *Journal of Computer-Mediated Communication*, *19*(4), 1024-1041.

Rainie, L. (2012). Social media and voting. *Pew Research Center Internet and American Life Project*.

Renninger, B. J. (2014). "Where I can be myself… where I can speak my mind": Networked counterpublics in a polymedia environment. *New Media & Society*. doi: 1461444814530095.

Rosenblat, A., Wikelius, K., Gangadharan, S. P., & Yu, C. (2014, October). Data & civil rights: Employment primer. In *Data & Civil Rights Conference*.

Sandvig, C. (2014). Corrupt personalization. *Social Media Collective, accessed July 5, 2014, http://socialmediacollective.org/2014/06/26/corrupt-personalization/; Jonathan Zittrain,"Facebook Could Decide*, 43-46.

Shirky, C. (2011). The political power of social media: Technology, the public sphere, and political change. *Foreign Affairs*, 28–41.

Shirky, C. (2015, March). The Value of digital data. Retrieved from http://www.cjr.org/analysis/the_value_of_digital_data.php

Sifry, M. (2014, October 31). Facebook wants you to vote on Tuesday. Here's how it messed with your feed in 2012. *Mother Jones*.

Simonite, T. (2015, January 9). *Robot journalist finds new work on Wall Street*. MIT Technology Review. Retrieved from http://www.technologyreview.com/news/533976/robot-journalist-finds-new-work-on-wall-street/

Singer, P., Flöck, F., Meinhart, C., Zeitfogel, E., & Strohmaier, M. (2014, April). Evolution of Reddit: From the front page of the internet to a self-referential community? In *Proceedings of the companion publication of the 23rd*

international conference on World wide web companion (pp. 517-522). International World Wide Web Conferences Steering Committee.

Srinivasan, S. (2014, April 14). *YouTube yearly costs for storage/networking – estimate*. Retrieved from https://sumanrs.wordpress.com/2012/04/14/youtube-yearly-costs-for-storagenetworking-estimate/

Stoddard, G. (2015). Popularity and quality in social news aggregators: A study of Reddit and Hacker News. *arXiv preprint arXiv:1501.07860*.

Stryker, C. (2012). *Hacking the Future*. Penguin.

Sweeney, L. (2013). Discrimination in online ad delivery. *Queue*, *11*(3), 10.

Terveen, L., Konstan, J. A., & Lampe, C. (2014). Study, build, repeat: Using online communities as a research platform. In *Ways of Knowing in HCI* (pp. 95–117). Springer.

Central Intelligence Agency. (2013). *The World Factbook* 2013-14. Retrieved from https://www.cia.gov/library/publications/the-world-factbook/index.html

Tufekci, Z. (2008). Grooming, gossip, Facebook and MySpace: What can we learn about these sites from those who won't assimilate? *Information, Communication & Society*, *11*(4), 544–564.

Tufekci, Z. (2014). Big questions for social media big data: Representativeness, validity and other methodological pitfalls. In ICWSM '14: *Proceedings of the 8th International AAAI Conference on Weblogs and Social Media*.

Tufekci, Z. (2014). Engineering the public: Big data, surveillance and computational politics. *First Monday*, *19*(7).

Tufekci, Z. (2014, August 14). *What Happens to Ferguson Affects #Ferguson: Net Neutrality, Algorithmic Filtering and Ferguson.* Retrieved from https://medium.com/message/ferguson-is-also-a-net-neutrality-issue-6d2f3db51eb0

Tufekci, Z. (2014, September 4). *Why Twitter should not algorithmically curate the timeline.* Retrieved from https://medium.com/message/the-algorithm-giveth-but-it-also-taketh-b7efad92bc1f

Walther, J. B. (2012). Affordances, effects, and technology errors. *Communication Yearbook*, *36*, 190-194.

Wang, H., Can, D., Kazemzadeh, A., Bar, F., & Narayanan, S. (2012, July). A system for real-time twitter sentiment analysis of 2012 us presidential election cycle. In *Proceedings of the ACL 2012 System Demonstrations* (pp. 115-120). Association for Computational Linguistics.

Washington, A. L., Parra, F., Thatcher, J. B., LePrevost, K., & Morar, D. (2013). What is the correlation between Twitter, polls and the popular vote in the 2012 Presidential Election?

Williams, O. (2015, March 25). *Facebook has officially declared it wants to own every single thing you do on the internet.* Retrieved from http://tnw.to/p3NMu

Xu, J., Lu, T. C., Compton, R., & Allen, D. (2014, June). Quantifying cross-platform engagement through large-scale user alignment. In *Proceedings of the 2014 ACM conference on Web science* (pp. 281-282). ACM.

York, J. (2012). Arab Digital Vanguard: How a Decade of Blogging Contributed to a Year of Revolution, The. *Geo. J. Int'l Aff.*, *13*, 33.

Zimmer, M. (2010, May 14). *Facebook's Zuckerberg: Having two identities for yourself is an example of a lack of integrity.* Retrieved from http://www.michaelzimmer.org/2010/05/14/facebooks-zuckerberg-having-two-identities-for-yourself-is-an-example-of-a-lack-of-integrity/

Zittrain, J. (2014, June 1). Facebook could decide an election without anyone ever finding out: The scary future of digital gerrymandering—and how to prevent it. *The New Republic.*

Zuckerman, E. (2014). Cute cats to the rescue? Participatory media and political expression.