

what went wrong:

A Design Oriented Response to Techno-Pessimism in Social Media

Sarp Gurakan

Harvard University

March 2019

Undergraduate Thesis for Joint Degree in Computer Science and Philosophy

Introduction

When the newspaper appeared in the 18th century, Enlightenment thinkers praised it as the necessary medium for the development of a public sphere in which individuals could engage in rational reflection and the advancement of society.¹ Soon after, Soren Kierkegaard castigated the very same medium for creating a place that allowed only for pure reflection, without any opportunity for action.² Following the global acclaim of the television for the audience size it is able reach, Martin Heidegger pessimistically noted on the impending loss of any understanding of the closeness of ideas or people.³ Every time early enthusiasts made grand claims on the possibilities some new medium would open up, critics equating the medium to a dystopia soon followed.

In a similar fashion, the first fanatics of the Internet imagined that its unprecedented ease of communication would eventually bring human beings closer together and allow for meaningful interactions between those who previously had no access to one another. In its early days, techno-optimists described the World Wide Web as the ultimate abolishment of barriers in global communication. Utopic euphemisms such as the unbounded "information highway" were used to articulate the beginning of a wholesome interconnectedness that was rivaled by no other age.⁴

Following its historical predecessors, the Internet soon met its techno-pessimistic critique. One of the first arguments against the Internet was that it would make all alternative forms of communication irrelevant and declare its exclusive reign as a "technopoly."⁵ Later, the question

¹ Habermas, Jurgen, Jürgen Habermas, and Thomas McCarthy. *The structural transformation of the public sphere: An inquiry into a category of bourgeois society*. MIT press, 1991.

² Dreyfus, Hubert L. *On the internet*. Routledge, 2013.

³ Heidegger, Martin. "The Thing." *Poetry, Language, Thought.*, Harper & Row, 2001.

⁴ Dreyfus, *OL*.

⁵ Postman, Neil. *Technopoly: The Surrender of Culture to Technology*. Vintage Books, 1993.

arose whether or not *Second Life*, a role-playing computer game, would replace the physical world that human beings inhabit.⁶ Today, techno-pessimism targets social media most fiercely. Claims against social media vary. Some suggest that it is not a platform of mass communication, but mass dissemination.⁷ Others claim that YouTube is creating a dystopia for the sake of higher ad revenue,⁸ and that Facebook has undermined democracy in an irreparable way.⁹ In fields of social anthropology, tech sociology and political science, social media constitutes a considerable part of the prominent discussion on technology and society.

There is no doubt that the harsh criticism social media faces, articulated by techno-pessimists, is insightful. These theoretical discussions are supported by empirical epidemics. In 2018 alone, Facebook has been the subject of public outcry for allowing foreign influences to impact US general elections, was cited by the UN as one of the actors responsible for the religious slaughter of Muslims in Myanmar and had its founder Mark Zuckerberg testify before the American congress for leaking close to 90 million users' data to a third party.¹⁰ Although Facebook seems to be the most scandal-ridden of the social media platforms, each has received its share of public outcries. Even during the first months of 2019, YouTube was partially blamed for the rise of anti-vaccination movements, which was ranked by the World Health Organization as one of the

⁶ Dreyfus, *OI*.

⁷ Gershon, Ilana. "Un-friend my heart: Facebook, promiscuity, and heartbreak in a neoliberal age." *Anthropological Quarterly* 84, no. 4 (2011): 865-894.

⁸ Tufekci, Zeynep. "We're building a dystopia just to make people click on ads." *TED: Ideas Worth Spreading*, Sept (2017).

⁹ Vaidhyanathan, Siva. *Antisocial media: How Facebook disconnects us and undermines democracy*. Oxford University Press, 2018.

¹⁰ Lapowsky, Issie. "The 21 (and Counting) Biggest Facebook Scandals of 2018." *Wired*. December 21, 2018. Accessed March 26, 2019. <https://www.wired.com/story/facebook-scandals-2018/>.

top 10 global health threats.¹¹ As to the amount of backlash social media platforms have faced supports, there is a definite need for a critique of social media.

However, the judgement passed by techno-pessimists misses a critical point of distinction between the aspects of social media that are inherent and those that are added by design. In other words, the condemnation of social media, while perceptive, is often misdirected at the medium itself. As a set of applications that exists within the World Wide Web, which itself exists as an application within the Internet, the architecture of social media is bound solely by the rules of communication of the Internet. The design of the Internet follows an "hourglass model," in which the layers are largely independent of one another, and the middle of the hourglass is intentionally designed narrowly.¹² This means that applications on the Internet are designed to be as independent as possible from the many layers of technology surrounding them. In terms of how social media platforms behave, little is due to the limitations of the concept of social media, while much is due to design choices made by the owners of each particular platform. Beyond its particular implementations, "social media" stands for the more general and flexible concept of globalized public communication through the Internet. Therefore, a criticism of *social media* as a concept is unproductive. Instead, what must be questioned are problematic design decisions that can be fixed with alternate design forms.

In other words, the question of social media is not a medium question, but a design question. In this instance, *design* does not simply refer to a set of aesthetic decisions, which is often how the word has been colloquially used and interpreted. Rather, in the technical sense of the word, *design* is a set of structural and architectural decisions. Anything that forms the frame around the content

¹¹ Wong, Julia Carrie. "How Facebook and YouTube Help Spread Anti-vaxxer Propaganda." The Guardian. February 01, 2019. Accessed March 26, 2019. <http://www.theguardian.com/media/2019/feb/01/facebook-youtube-anti-vaccination-misinformation-social-media>.

¹² Zittrain, Jonathan. *The Future of the Internet--and How to Stop It*. Yale University Press, 2009.

of the platform is considered design. Through intended consequence and unintended side effects, the design of a platform nudges the user towards certain behaviors, encourages some types of interactions over others, and has an undeniable influence over the totality of ways in which consumers use the platform.

Design not only refers to what is made possible and what is not. Just as significantly, what is made accessible and what is left inconspicuous belongs to the theory of design as well. Even though a technologically literate and skilled user may be able to use the full spectrum of the platform's possibilities, most users will limit themselves to the uses that they see as clearly defined and immediately beneficial. The 80/20 rule, applied to human-computer interaction, declares that around 80 percent of users will use around 20 percent of the functionality of any given platform—the most exposed and basic features.¹³ Therefore, the wide capabilities of social media are always categorized and prioritized in certain ways. Instead of condemning the limitations of online public communication, a fruitful discussion must investigate social media through the lens of design theory.

In the following pages, this work discusses the central critique of social media, some of which has already been alluded to. It attempts to rescue the medium itself from condemnation, instead redirecting the criticism at particular design decisions that have given rise to the problems articulated. Admittedly, the belief that the medium is worth saving only follows from an implicit conviction that public communication is valuable and deserves advancement. The discussion of the legitimacy of the ideal of global communication is left to the field of moral philosophy. Instead,

¹³ "The Pareto Principle and Your User Experience Work." The Interaction Design Foundation. Accessed March 26, 2019. <http://www.interaction-design.org/literature/article/the-pareto-principle-and-your-user-experience-work>.

this work focuses on the claim that this ideal of global public communication can be sheltered from the critique that its current realizations should face instead.

To assure the validity of this claim, alternative design choices that would alleviate the critics' concerns are put forth. Specifically, a wide set of arguments are categorized into three, according to the design features from which they originate—notification settings, structure of posts, and engagement-based optimization algorithms. For each of these categories, the claims made against it is explored and evidenced. Thereafter, its connection to the given design feature is demonstrated, and suggestions for alternative design features are provided. Later, counter arguments are anticipated in three groups: system-wide arguments, general arguments that pertain to multiple claims, and singular arguments. Following a response to these challenges, a discussion of the wider implications of this thesis is offered.

Medium and Design

Notification Settings

Sherry Turkle, one of the leading tech sociologists of today, articulates one of the main problems exacerbated by social media as such: "We are confused about when we are alone and when we are together."¹⁴ Despite its simplicity, this claim is not a revision of the all-too-frequently uttered lamentation that people are on their phones too much. Turkle's claim is also not one about how recent technological developments no longer allow us to communicate in more personal ways.

Actually, the claim is not about devices at all. Rather, Turkle suggests that while incorporating networked communication into our lives, we have created virtual selves for whom many constraints of the physical selves do not apply. For example, she says of the virtual self, "It can absent itself from its physical surroundings [...], it can experience the physical and virtual in near simultaneity. And it is able to make more time by multitasking."¹⁵ Moving into a different realm when one is dissatisfied with the physical, performing multiple tasks at the same time, being in to places at once—all these properties seem to be improvements to the human experience at a first glance. However, the problem, according to Turkle, stems from the confusion the user feels between the physical and the virtual self.

Turkle hammers this point home when she grants that "we have always found ways to escape from ourselves, neither the desire nor the possibility is new with the Internet."¹⁶ However, she claims, the difference that makes the Internet a worthwhile case study is its capacity to weave together the virtual and the physical and allow a simultaneous access to both, a "life mix" that is

¹⁴ Turkle, Sherry. *Alone Together: Why We Expect More from Technology and Less from Each Other*. Basic Books, 2017. 329.

¹⁵ Ibid, 155.

¹⁶ Ibid, 160.

co-constituted by one's life online and offline.¹⁷ This capability does not inherently harm its users, but Turkle's findings suggest that users now partition their attention to these two lives in a sense of "continual copresence."¹⁸ This misuse transforms the capabilities and advantages of networked communication into social and personal deficiencies. The ability to experience two worlds simultaneously harms our social interactions in physical spheres, as we begin to think of people as "pausable."¹⁹ Similarly, the ability to communicate instantly and at any time ironically hinders communication, because "when media are always there, waiting to be wanted, we lose a sense of choosing to communicate."²⁰ The ability to be connected to a wider world is transfigured into a feeling of "always feeling behind", or a fear of missing out, abbreviated by Internet culture as "FOMO."²¹ One study defines FOMO as "the pervasive apprehension that others might be having rewarding experiences from which one is absent" and finds that the self-reported feeling is strongly correlated with the user's level of social media engagement.²²

The main pitfall we must avoid while reading Turkle's findings is to regard the negative consequences of networked communication as inescapable conclusions that we must accept from the outset. The limitations that are native to the system versus the problematic design choices are clear in Turkle's analysis. The set of design choices that are relevant in the users' misuse into a "life mix" is ones regarding notification settings.

¹⁷ Ibid.

¹⁸ Ibid, 161.

¹⁹ Ibid.

²⁰ Ibid, 163.

²¹ Ibid, 164.

²² Przybylski, Andrew K., et al. "Motivational, emotional, and behavioral correlates of fear of missing out." *Computers in Human Behavior* 29.4 (2013): 1841-1848.

If the main gate that connects the physical and virtual worlds is the phone screen, then the notification icon is the knock on the gate. Especially as Turkle is worried about the "[movement] into the virtual with fluidity and on the go," the *Red Alert Notification*²³ is the symbol that ignites this transition. The notification icon serves as an indicator that some activity that is pertinent to the user is taking place. This occurrence is possibly personally addressed to her, and the icon signifies that the activity has momentarily taken place. The alert is likely to create a sense of urgency, especially since it arrives in real time and is capable of triggering the aforementioned fear of missing out.²⁴ As a bright red bubble that places itself in the app icon and makes itself readily visible in the user's home screen, it invites the user to engage with the platform. Notifications have the power to alert the user to relevant activity, but they also run the risk of alluring the user into the platform even when it is not in her best interest.

The multi-faceted positioning of notifications gives significance to how notification settings are composed. An investigation into the notification settings of social media platforms quickly reveals a point of deficiency.

The first declaration in the notification settings page of Facebook declares, "You can't turn off notifications entirely, but you can choose what you're notified about and how you're notified."²⁵ Reasonably, Facebook expects that users are interested in being notified about certain activity, since this is the general promise of networked communication. However, determining how and when a user wished to be notified turns out to require considerable effort from the user.

²³ "Notification Basics & Settings | Facebook Help Center." Facebook. Accessed March 26, 2019. <https://www.facebook.com/help/327994277286267>.

²⁴ Turkle, *AT*, 164.

²⁵ Facebook. Accessed March 26, 2019. https://www.facebook.com/settings?tab=notifications&ion=on_facebook&view.

For any Facebook user, all notification settings are on by default, meaning that unless the user changes the given notification settings, he will be notified by every category of activity that Facebook has curated. As the statement reads: "You'll see every notification on Facebook, but you can turn off notifications about specific posts as you view them."²⁶ Unless the user is satisfied with an all-inclusive notification system, he has to turn off notifications on a case by case basis—a completely *opt-out* policy. In order to even begin the process of turning off certain notifications, the user must have a certain level of awareness and capability. Since Facebook does not force the users to determine their own notification settings during the sign-up process, users are more likely to operate on the default settings than not.

Even if a user embarks on the journey of customizing the settings, progress is unexpectedly slow, given the number of notification settings. For example, many third-party applications have incorporated a "Continue with Facebook"²⁷ button as an alternative to signing in separately, which automatically turns on notifications from each of these apps. Without a way of disabling third-party application notifications platform-wide, the user has to go through each application and turn it off individually.

Facebook's frequent application updates make the process of customizing notification settings a regular endeavor. New types of notifications are introduced with each update, and these notifications are also turned on by default. As *Android Central* notes: "Facebook likes to do a lot of "opt out" rather than "opt in" changes on its app updates."²⁸ Every time the Facebook app is updated, the burden of finding and customizing the new settings fall on the user.

²⁶ Ibid.

²⁷ "Login Button - Facebook Login - Documentation." Facebook for Developers. Accessed March 26, 2019. <https://developers.facebook.com/docs/facebook-login/web/login-button/>.

²⁸ Martonik, Andrew, and Andrew Martonik. "How to Turn off the New Persistent Facebook Notification." *Android Central*. May 10, 2013. Accessed March 26, 2019. <https://www.androidcentral.com/how-turn-new-persistent-facebook-notification>.

In certain cases, the newly introduced notification settings near the limit of what can reasonably be considered a "notification." For example, *Android Central* explains the addition of a particularly intrusive feature, "a persistent 'ongoing' notification that lives in your notification pull-down."²⁹ This notification does not signify any particular activity, and it reappears once the user removes it from the smartphone screen. Although this may have been a bug, it was never reported as such, and shows the greedy nature of Facebook notifications.

This shortcoming is exacerbated by the fact that Facebook often does not document particular changes in the update description at all.³⁰ Occurring around once to twice a week, Facebook updates its app with the same description for every update. The generic statement in each update reads: "Thanks for using Facebook! To make our app better for you, we bring updates to the App Store regularly. Every update of our Facebook app includes improvements for speed and reliability. As new features become available, we'll highlight those for you in the app."³¹ The lack of documentation makes it especially difficult for users to customize their notification settings.

Opt-out and frequently expanding notification settings may push the users to partake in the *life-mix* even when it is not their intention. Users' engagement with a social media platform may begin with the intention of checking some activity, and it may continue beyond the specific activity inadvertently. This accidental engagement is partially due to the inviting function of the notification icon.

An example of this phenomenon in the real world is the promotional sign outside a retail store. Inside the store, the items on sale often neighbor those that are full-priced. The discerning

²⁹ Nickinson, Phil, and Phil Nickinson. "Facebook Quietly Adds Facebook Home Support (and Unofficial Support) for Galaxy S4, HTC One, Others." *Android Central*. May 10, 2013. Accessed March 26, 2019. <https://www.androidcentral.com/facebook-quietly-turns-support-and-unofficial-support-galaxy-s4-htc-one-others>.

³⁰ Ibid.

³¹ Babulous. "Why I Kicked Facebook off My Phone." *Hacker Noon*. April 15, 2017. Accessed March 26, 2019. <https://hackernoon.com/why-i-kicked-facebook-off-my-phone-a254c5b410bc>.

customer knows that once he enters the store, his attention may unintendedly move beyond the items on sale—to be grabbed by an item from the new season. Once the customer's initial motivation to check out the sales is overcome, he is as likely as other customers to purchase full-priced items.

To combat the issue of unintended visits to social media, an alternative platform can allow the user to make broader changes in push notifications. A "conservative" meta-configuration can configure the settings to be completely off, so that the user can selectively turn on the notifications that are of interest. The fully opt-out settings can also remain under another meta-configuration like "permissive," so that the user that wishes to engage with the platform maximally can continue to do so. Finally, a "power user" meta-configuration can allow the more technically capable and discerning users access to more advanced settings. Requiring the user to decide among these three options during the sign-up process guarantees that users are not stuck with notification settings that are not beneficial for them.

Admittedly, there is an expansive literature on the tradeoff between opt-in and opt-out strategies. The complications introduced by the aforementioned recommendations are discussed in the next chapter in the form of counter arguments.

Invariant Post Structures

In a damning criticism, Siva Vaidhyanathan's *Antisocial Media* evidently argues that "the problem with Facebook is Facebook," meaning that the platform is inherently problematic and cannot be saved by any improvements.³² On the other hand, Vaidhyanathan's judgments on Facebook, translatable to other social media platforms as well, are mainly pointed at a single design choice.

³² Vaidhyanathan, *AM*, 1.

Vaidhyanathan's first claim is that Facebook makes it a challenge to distinguish between high-quality and low-quality sources and causes the users to conflate what is entertainment with what is more vital. He evidences this claim with the observation that in Facebook, distinguishing between different types of content requires careful attention.³³ Whether it is a news article, a friend's vacation photos, or a post announcing a loved one's death, all posts have the same basic structure: A description of the activity type, a text, an optional visual component and the reactions bar. Furthermore, the main form of travel inside Facebook is "scrolling through"³⁴, meaning that there is no distinct movement between individual posts. Indeed, the same observations hold for the other platforms as well—Instagram switches the placement of the visual component and the text; Snapchat works with clicks instead of scrolls, and Twitter emphasizes the text over the visual. Besides these small differences, the invariant structure between posts is a constant in all of these platforms.

The field of pedagogy has ample evidence to suggest that humans are not context-independent learners. We make sense of information with context systematically better than without context.³⁵ The stronger of a connection there is between pieces of information presented sequentially, the more unified our experience is. On the other hand, the presentation of content in all the aforementioned social media platforms is through a whirlwind of discontinuous posts that are not curated by relevance and cover a wide range of material and tone. Without an understanding of a gradient, the user is left without the possibility a unified experience after a visit to a social media platform. The user, without the guidance of a unifying ground, ends up scrolling through

³³ Ibid, 4.

³⁴ Ibid, 5.

³⁵ Dreyfus, *OI*, 56.

the posts purposelessly, and retains the bits and pieces of information that catch her eye in the process.³⁶

Unfortunately, the types of posts that catch one's eye tend to be the "sensationalist" ones.³⁷ Posts that include radical language, refer to extreme events or evoke strong emotions stand out from the rest of the feed. As posters become more and more aware of this phenomenon, they are incentivized to create posts of this sort, instead of curating reflective and moderate content. As posters move towards brighter spectacles, the spectators lose sight of what is pure entertainment, what is personally intimate, and what is politically significant. The identical structure of all posts disallows the possibility of differentiation by any method other than radicalization.

Furthermore, due to practically unending home feeds, the user has virtually infinite content that she can scroll through. There is little incentive to contemplate any given post, since the FOMO-like urge to see as much as possible is at play.³⁸ The user may unreflectively share an article without reading it, or like a post only by reading the first sentence. With so much content that is difficult to categorize, and so little time for each post, the user rarely comes away with a meaningful digestion of the information she has just observed. Vaidhyathan puts it thus: "On Facebook babies and puppies run in the same column as serious personal appeals for financial help with medical care, advertisements for and against political candidates, bogus claims against science, and appeals to racism and violence."³⁹

Another problem caused by the limited attention is afforded to each post is the disguise of sponsored ads as regular posts. In Facebook, sponsored content has the same format as genuine posts, with the sole difference of a sub-text that reads "sponsored." To blend in with regular posts,

³⁶ Vaidhyathan, *AM*, 35.

³⁷ *Ibid*, 44.

³⁸ *Ibid*.

³⁹ *Ibid*, 17.

ads often use material that do not reveal themselves to be ads during the initial engagement. For example, an ad for the body-building app *V Shred* presents itself as a free survey about healthy living, which reveals itself to be an ad only at the end of the survey.⁴⁰ Similarly, Instagram ads are photos posted in the same style as other posts, although with an option to click on them. Snapchat stories include sponsored videos that often begin as regular stories that turn out to be ads.

It is common knowledge that ads are meant to manipulate their viewers, and they are met with suspicion⁴¹. Therefore, this marketing strategy is a useful tactic to get users to retain the given information, since they do not initially recognize it as part of an ad. In addition, because ads are targeted, the content is often relevant to the users' interests—meaning they are likely to believe the content to be genuine. As a recent Stanford University study observed, "more than 80 percent [of participating students] believed a native ad, identified with the words 'sponsored content,' was a real news story."⁴² It is not surprising that one might not be able to discern between sponsored content and regular content during a quick scroll-through.

Similarly, a ProPublica study on the US 2018 midterm elections discovered political ads that misrepresented themselves as news organizations, such as "Ohio Newswire" and "Breaking News Texas."⁴³ Once again, this is a sensible strategy for advertisers, and it makes sense for social media platforms to take advantage of it. Nonetheless, it is undeniable that this strategy of advertising is in many ways detrimental to user experience.

⁴⁰ V Shred's Facebook Page, Accessed March 26, 2019.
<https://www.facebook.com/watch/?v=2267436653579031>

⁴¹ Tufekci, *Dystopia*.

⁴² Wineburg, Sam, Sarah McGrew, Joel Breakstone, and Teresa Ortega. "Evaluating information: The cornerstone of civic online reasoning." *Stanford Digital Repository*. Retrieved January 8 (2016): 2018.

⁴³ Merrill, Jeremy B. "What We Learned from Collecting 100,000 Targeted Facebook Ads." ProPublica. March 09, 2019. Accessed March 26, 2019. <https://www.propublica.org/article/facebook-political-ad-collector-targeted-ads-what-we-learned>.

Vaidhyathan's description of the condemnation fully targets a single design choice, although his initial claim contradicts this fact. Invariant post structures are not a given for social media, but an intentional decision. Indeed, when we bifurcate social media platforms according to how they are organized, we encounter an entire set of platforms that separate different types of content. Platforms such as Reddit and Quora have threading features that basically organize posts by subject category and require the user to post the content together with other pieces of content that fall under the same type or tone. With different moderation rules, each thread is afforded its own norms and community. Social media platforms that are organized by the relationships between users can adopt similar threading features, so that both the trivial and the sober can have their own places.

A seemingly similar yet fundamentally separate criticism that falls under the same design feature is explored in *On the Internet*. In this book, Hubert Dreyfus warns us against the levelling effect of organizing a platform without a concern for inter-post context.⁴⁴ Due to the undifferentiation of the content of each post, the user cannot help but view all posts with the same mentality; she automatically equalizes them in value. Since a trivial post, such as a puppy photo, cannot be elevated to the status of the significant, all posts are viewed in the most trivial light, from a place of "detached reflection."⁴⁵

In Dreyfus' terminology, this is a type of virtual nihilism in which no matter how essential the content of a post is, it cannot move the user to engage with the material in a way that demands effort. In other words, the posts are not likely to push the users to take action that extends outside of the platform, since the ones that have that capability are drowned by the ones that are

⁴⁴ Dreyfus, *OI*, 75.

⁴⁵ Ibid, 72.

commonplace. In his words: "Nothing is too trivial to be included. Nothing is important that it demands a special place."⁴⁶ Social media ends up being a platform that only consumes attention and time, without granting the possibility of production of anything meaningful in the physical world. The user is reduced to the lowest common denominator in which he can make sense of all of the information. However, this common denominator, by absolute inclusion, becomes infinitesimal. Therefore, the reader's grounds are absolutely abstract, such that he "[cannot have] an essential engagement in anything."⁴⁷

This issue of levelling stems from most platforms' lack of a capability to differentiate between user groups. Whereas our physical lives are governed by different sets of norms and rules depending on the social circle we are in at a given moment, social media platforms by default conflate all circles into one home feed, and only some of them allow for a different grouping at all. Therefore, we cannot dictate how we would like to interact with a certain group of people; we must cater to the entire mass following.

Social media platforms are designed in a way to encourage the dissipation and visibility of one's material. When the user posts, the tendency (and perhaps the purpose) of the platform is not to keep the post private but make it available to the public. Even though there are a few options for the user to target a specific group of people in the post, they are not made to be the primary use case scenarios of the platform. For example, Facebook allows for the user to create a group for "best friends" and target a post towards this individual group, but the default is still to make the post public to all of one's connections. Similarly, Instagram allows the user to make his profile 'private' such that he has to approve every user that gets access to the profile, but the default setting is that all profiles are open to the public.

⁴⁶ Ibid, 79.

⁴⁷ Ibid, 76.

Furthermore, both platforms have a "suggested friends" bar that appears at the center of the feed. Thus, the user is continuously encouraged to make connections with new people. There is also no way to limit the sharing of one's posts—any given post the poster posts can be shared by any of its recipient, making it virtually impossible for the poster to get a hang of the users that will eventually read the post. One way in which we can see that this has bothered some users through the evolution of the concept of a "Finsta," a secondary account that the user intentionally keeps secretive so that he can share more personal material with a selected group.

One way to show the inadequacies of having to target all of one's following in each post is using Sartre's explanation of an extreme anxiety that comes from the feeling of being watched by an unknown other, applied to social media. Using the example of the Sartrian concept of "the look,"⁴⁸ the poster who is not comfortable with the potential audience of her content experiences a peculiar anxiety of being watched by "the other."⁴⁹ The other is a particular individual that observes her judgmentally. Recognizing the presence of the other, the poster can no longer experience herself as a true subject. As Sartre puts it: "By the mere appearance of the Other, I am put in the position of passing judgment on myself as an object, for it is as an object that I appear to the Other."⁵⁰ Even if the individual is not correct about her being watched by the foreign Other, the mere appearance of being watched causes the loss of subjectivity. This anxiety causes a fundamental change in the subject's experience of herself; she begins to see herself only in relation to the other.

Once again, it is not native to the concept of social media that the poster must feel such an anxiety. The anxiety is felt only when the user feels that the post might reach users for whom it

⁴⁸ Sartre, Jean-Paul. "The Look". *Being and nothingness*. Open Road Media, 2012.

⁴⁹ Ibid.

⁵⁰ Ibid.

was not intended. Although this possibility cannot simply be removed, it can be alleviated by a similar grouping feature.

Just like a platform can categorize content by subject, it can also create a distinction between mass following vs following by group. Making "following by group"—or navigating the social media platform in the framework of different groups—can fix the problem of the stranger's look and the alienation it creates. Different levels of significance and intimacy can be separated by design. As Vaidhyathan notes in the aforementioned piece: "Different forms of friendship have distinct layers and values embedded in them and operate by different norms."⁵¹ Therefore, the conflation of different layers causes a reduction to lowest common denominator, a denial of strong forms of engagement, and an anxiety caused by the possibility of the invasion of one's perceived privacy. Concretely separated communities with differentiated and explicit norms to which each post alleviates this problem.

Optimization Algorithms

When there is practically infinite content, curation consumes content. What we experience is no longer a function of what content is available in the platform, but how the platform curates, orders and presents its content. With this in mind, the question of curation becomes the subject matter of the critique of Zeynep Tufekci, as well as Vaidhyathan.

Vaidhyathan notes that Facebook mainly ranks the posts on the news page using the criterion of "engagement," which is calculated by the number of reactions, comments and shares.⁵² This mechanism builds a positive feedback loop for posts that get early recognition, and a negative feedback loop for the posts that do not get attention early on. If a post gets some engagement when it is first posted, it is more likely to resurface on other users' home pages. Given more visibility,

⁵¹ Vaidhyathan, *AM*, 47.

⁵² *Ibid*, 6.

the post is more likely to get even more attention, in a positive cycle. Similarly, posts with little engagement early on are less likely to appear on other users' home pages, which makes them less likely to get engagement later on. In other words, a post survives as long as it gets early attention. Otherwise, it dies quickly.

From the spectators' perspective, this has the effect of further incentivizing radical content. Vaidhyathan talks about this issue in terms of "fake news" and "filter bubbles," but the problem translates to all types of content.⁵³ Although this does not mean that more nuanced posts do not make it in users' home pages, it means they appear less frequently than their more provocative counterparts. Vaidhyathan says: "most inflammatory material will travel the farthest and the fastest. Sober, measured accounts of the world [will not]."⁵⁴

Similarly, from the posters' perspective, reaching more of their social circles requires that they cooperate with this curation algorithm. *Click-baits*—posts that grab the users' attention quickly and require low levels of engagement—gain popularity. As the posters' patterns of posting end up validating the spectators' conception of social media, the cycle of loss of nuance, appeal to lowest common denominator and increasing polarization continues.

Another factor in curating one's home page is the user's previous activity. Platforms decide on what content to show by calculating a predictive score of each item⁵⁵ that anticipates how likely the user is to engage with the post given her history. In a basic understanding, if the user has engaged with a certain kind of post in the past, she is more likely to engage with similar kinds of posts in the future. As Vaidhyathan notes, this contributes to the same politically polarizing consequences, such as the formation of filter bubbles: "[An] intellectual isolation [occurs] when

⁵³ Ibid, 15.

⁵⁴ Ibid.

⁵⁵ Ibid.

websites make use of algorithms to selectively assume the information a user would want to see, and then give information to the user according to this assumption."⁵⁶ Instead of being in a global conversation, users end up inhabiting disjointed circles that repetitively create the same type of content.

Another worry concerns the *Up Next* column that makes use of Youtube's recommendation algorithm. Tufekci claims that looking at a user's watch history, YouTube constructs a personality type for the user to show him videos that "people like him" would watch.⁵⁷ This results in an ever-increasing extremism in the types of videos YouTube recommends, not just in political content but in all types of material. If the user watches a video about vegetarianism, he gets a video recommendation about veganism— As Tufekci puts it: "It's like you're never hardcore enough for YouTube."⁵⁸ The algorithm pigeonholes the user to a certain type of personality. Through the mere exposure effect, it is even possible that the user ends up fitting his archetype more and more by engaging with his recommended videos, ironically fulfilling his YouTube crafted destiny.

Targeted ads use the same recommendation system that groups users into different "types of people." The aforementioned ProPublica study shows that Facebook has a "lookalike audience"⁵⁹ ad-targeting feature, which allows advertisers to use the groupings that the Facebook algorithm makes for curation purposes. The study displays that in the 2018 US mid-term elections, more than 70 percent of all political ads targeted one side of the political spectrum at least twice as much as the other. Almost three quarters of political ads were seen by voters who were identified as possibly being convinced by those ads, and less than 20 percent of ads made it to people from

⁵⁶ "What Is a Filter Bubble? - Definition from Techopedia." Techopedia.com. Accessed March 26, 2019. <https://www.techopedia.com/definition/28556/filter-bubble>.

⁵⁷ Tufekci, *Dystopia*.

⁵⁸ Ibid.

⁵⁹ Merrill, Jeremy B. *Targeted Facebook Ads*.

both sides of the spectrum close to bilaterally.⁶⁰ This means that for most parties, it was impossible to contest the facts of an ad from the other side of the spectrum, simply because they never even received those ads. The possibility of social media as a public sphere significantly deteriorates when users do not have common grounds through which to engage in public discussion.

These issues all find realization in the optimization algorithms employed by the platform. optimization of an algorithm is human-made and therefore subject to change. There is no method of general optimization for an algorithm; optimization is always geared at a certain goal. An optimization algorithm seeks to minimize a given definition of a cost, but it is up to the programmer to determine what the parameter to be reduced is. Then, the cost function attempts to minimize the difference between the expected value of a parameter and the empirical value of that parameter. Perhaps pure engagement is a convenient parameter to optimize, but it is not necessarily the correct parameter, and it is decisively not a unique parameter. Thus, by picking cost functions that optimize more complex notions of user satisfaction than merely engagement, an alternative social media platform could be protected from these types of criticism.

Furthermore, social media platforms recently began updating and developing their cost functions to mediate some of the very problems described. In January 2019, YouTube has released a statement explaining how they have been combatting videos with *click-bait*:

You might remember that a few years ago, viewers were getting frustrated with clickbaity videos with misleading titles and descriptions (“You won’t believe what happens next!”). We responded by updating our system to focus on viewer satisfaction instead of views, including measuring likes, dislikes, surveys, and time well spent, all while recommending clickbait videos less often.⁶¹

⁶⁰ Ibid.

⁶¹ "Continuing Our Work to Improve Recommendations on YouTube." Official YouTube Blog. January 25, 2019. Accessed March 26, 2019. <https://youtube.googleblog.com/2019/01/continuing-our-work-to-improve.html>.

Admittedly, given that Youtube's algorithm is proprietary and well-protected, it is not an easy task to examine the accuracy of this claim or the nature of the update. Nonetheless, the declaration serves to show an understanding from Youtube's part that simple goals such as maximum views do not result in the best user experience—cost functions need to be improved to combat issues of radical content and better approximate the complex criterion of "viewer satisfaction."⁶²

In the end, the many blows social media takes from critiques are astute and necessary for the betterment of the platforms, but they must be viewed as design problems so that they can yield design solutions. Notification settings, post structures and optimization algorithms are specific implementations of features that are subject to radical change; none of the judgments uttered by the critics are unavoidably caused by the limitations of public communication over the Internet.

⁶² Ibid.

Challenges

Incentives

The main system-wide counter argument that advocates of change in social media face is that any alternative conception of a platform is unrealistic, since companies always make decisions that put their monetary interests ahead of user satisfaction. Thus, this criticism undermines productive dialog as to how a system can be improved by unconditionally rejecting the possibility of change.

Tufekci, in her aptly named article named *Yes, Big Platforms Could Change Their Business Models*, simulates this argument with the following lines:

There's simply no plausible alternative, the platforms say. People will never pay to use platforms, we are told. Plus, dissidents and activists in the developing world rely on these free services to get their word out. How can we abandon them? And anyway, the platforms say, we can't provide the fundamental features that our users value without all this data collection. It's simply too late to change.⁶³

Vaidhyathan concludes his analysis on a similarly pessimistic note. Claiming that condemning Facebook's practices is futile, he utters: "Facebook itself has no incentives to reform."⁶⁴ In other words, the general claim is that companies won't change simply because their customers demand it.

In reality, the question of incentives does not have an all-inclusive answer—what a company considers as incentive heavily depends on who it sees as its customer. Especially in ad-based platforms, a determination of who the users are is a complicated question. For example, some may claim that these platforms are designed for the benefit of the advertisers, and therefore user satisfaction is not a priority. However, in this case, platforms must risk the possibility of losing

⁶³ Tufekci, Zeynep. "Yes, Big Platforms Could Change Their Business Models." *Wired*. December 17, 2018. Accessed March 26, 2019. <https://www.wired.com/story/big-platforms-could-change-business-models/>.

⁶⁴ Vaidhyathan, *AM*, 17.

large number of users over time. Similarly, if platforms were to see only their users as customers, then advertisers may migrate to a platform that benefits them more. Therefore, even without access to the strategic development decisions that go into any ad-based platform, a reasonable assumption can be made that platforms should consider both users and advertisers as their customers. Hence, the question of business model is not a stand-alone; it is coupled with the question of customer base. In an interview, Mark Zuckerberg summarizes the main reason for this as such:

I think one of the key principles is that we're trying to run this company for the long term. And I think that people think that a lot of things that— if you were just trying to optimize the profits for next quarter or something like that, you might want to do things that people might like in the near term, but over the long term will come to resent. But if you actually care about [...] building the company for the long term, I think you're just much more aligned [with customers] than people often think companies are.⁶⁵

Of course, Vaidhyathan or any techno-pessimist can point out that simple utterances are not binding. However, empirical evidence shows that customer satisfaction is, indeed, binding.

An example of radical change in business models is the rise of what is called the "fast-casual"⁶⁶ food industry, and the fast adaptation of the largest fast-food chains to join the trend. In the last decade, the fast food industry saw the rise of a health-conscious customer base and had to change its practices drastically. Following the demands of their customer bases, McDonald's removed certain artificial preservatives from its menu; Taco Bell made a promise to become one

⁶⁵ Constine, Josh. "Highlights & Transcript from Zuckerberg's 20K-word Ethics Talk." TechCrunch. February 20, 2019. Accessed March 26, 2019. <https://techcrunch.com/2019/02/20/zuckerberg-harvard-zittrain/>.

⁶⁶ Carman, Tim. "Fast Casual Nation: The Movement That Has Changed How America Eats." The Washington Post. August 29, 2017. Accessed March 26, 2019. https://www.washingtonpost.com/lifestyle/food/fast-casual-nation-the-movement-thats-changing-how-we-eat/2017/08/28/23f6d710-86c5-11e7-961d-2f373b3977ee_story.html?noredirect=on&utm_term=.289b1ab110f1.

of the healthiest fast-food chains;⁶⁷ and the whole industry reduced the average calories of a menu item by 12% in one year.⁶⁸

The early symptoms of a similar migration are observable in the field of social media. Studies from CMU and Stanford—along with many others— showcase the significantly negative social and psychological effects of social media as it is experienced today.⁶⁹ It is not unreasonable to assume that a platform that allows its users' social lives to flourish as well as entertain them will more likely be adopted in the long run, and that the public will demand more from platforms eventually. The fast food industry, after decades of making food more and more chemically-developed and unhealthy, finally went through a tremendous change as the increase in diabetes rates finally caught the public eye.⁷⁰

Indeed, social media platforms have already started to see this backlash from users and shareholders. Facebook ended a scandal-ridden 2018 with a 20% drop in its stock price, the lowest it has been in nearly two years.⁷¹ Therefore, an alternative design is not only realistic, but perhaps even imminent. How a platform emerges need not govern how it develops. As Tufekci reminds us, seatbelts became mandatory in the US many decades after cars were popularized and were an indispensable part of life, following much controversy that argued that it was preposterous to demand that car manufacturers design all cars with such a major invariant.⁷²

⁶⁷ Turkle, *AT*.

⁶⁸ Bleich, Sara N., Julia A. Wolfson, and Marian P. Jarlenski. "Calorie changes in chain restaurant menu items: implications for obesity and evaluations of menu labeling." *American journal of preventive medicine* 48, no. 1 (2015): 70-75.

⁶⁹ Dreyfus, *OI*.

⁷⁰ Bleich, *Calorie Changes*.

⁷¹ Rodriguez, Salvador. "Here Are the Scandals and Other Incidents That Have Sent Facebook's Share Price Tanking in 2018." CNBC. November 21, 2018. Accessed March 26, 2019. <https://www.cnbc.com/2018/11/20/facebook-scandals-in-2018-effect-on-stock.html>.

⁷² Tufekci, *Big Platforms*.

Legitimacy of Radical Content

Through numerous arguments, the previous chapter has argued that social media platforms inadvertently favor and encourage radical content, while discouraging nuanced or calculated content. Two design choices have been displayed as responsible for this problem: The invariant structure of posts that causes users to give brief and shallow attention to any given post, and the cost functions that optimize for engagement and encourage users to evoke strong emotional responses their audiences. A possible counter argument that trivializes this claim is that the dissipation of radical content is not a legitimate problem.

A piece of evidence for this claim is that only a small portion of social media posts include radical and extreme material. Even though this material is covered in media and condemned by thinkers widely, one might suggest that it is these very authors that give undeserved attention to these outliers, thus granting them the popularity they seek. In this way, one may claim that radical posts are only as big as a problem as critics make of them. If left unattended to, they might be contained in a small area of the Internet without disseminating to the wider network.

Although this account is optimistic and convenient, it is nonetheless contrary to evidence. Even if the number of radical posts is small compared to moderate posts, they systematically pique users' interest. In the aforementioned interview, Mark Zuckerberg explains:

One of the findings that has been quite interesting is, [...] there's this question about whether social media in general [...] makes it so that sensationalist content gets the most distribution. [...] And what we found is that generally within whatever rules you set up [for what content is allowed], as content approaches the line of what is allowed, it often gets more distribution.⁷³

In other words, no matter how the content regulation rules are set up, users tend to engage with posts that are on the edge of the regulations more frequently. From Zuckerberg's perspective, this

⁷³ Constine, Josh, *Zuckerberg Highlights*.

is not a property of the system, but a property of human beings.⁷⁴ Indeed, there is a vast body of research that supports the claim that humans are attracted to the sensationalist and radical content. For example, a study that examines front pages stories of newspapers from eight countries published between the years 1700-2001 shows that the "sensational news" have made the front cover significantly more frequently than others.⁷⁵ Studies also show that attraction to gossip, scandals, and other types of content that create emotional response has been invariant throughout human history.⁷⁶ Some even suggest that gossiping about the scandalous is an evolutionary trait that humans have developed to enhance social bonding through dialog.⁷⁷ Therefore, in the case of human communication generally, and of social media content specifically, content that sparks emotional response is always more likely to get attention.

Furthermore, the prevalence of radical content is exacerbated by the network properties of social media. As a network, social media platforms connect all of its nodes to one another, resulting in a non-linear distribution that is called the "Network Effect."⁷⁸ Metcalfe's Law, which has been used to explain the network effect in many areas, states that the growth of an item in a network is in the order of the square of the number of nodes in the network that use that item.⁷⁹ In other words, networks make dissipation more efficient.

The application of Metcalfe's Law to social media demonstrates that discrepancies in popularity are heightened. This application explains the concept of "going viral"—If a post successfully beings an initial dissipation, it is more likely to spread like a virus. Although this is

⁷⁴ Ibid.

⁷⁵ Davis, Hank, and S. Lyndsay McLeod. "Why humans value sensational news: An evolutionary perspective." *Evolution and Human Behavior* 24, no. 3 (2003): 208-216.

⁷⁶ McAndrew, Frank T. "Can gossip be good?." *Scientific American Mind* 19, no. 5 (2008): 26-33.

⁷⁷ Ibid.

⁷⁸ Hendler, James, and Jennifer Golbeck. "Metcalfe's law, Web 2.0, and the Semantic Web." *Web Semantics: Science, Services and Agents on the World Wide Web* 6, no. 1 (2008): 14-20.

⁷⁹ Ibid.

generally a feature of networks, it adds further stress to the system in the case of radical content. Even if the sensationalist posts are low in number produced, their eventual dissipation is aggravated due to human factors and the network effect.

Possibility of Fair Algorithms

A challenge that targets the call to advance optimization algorithms is a technical argument that "fairness" is not a well-defined notion and that its pursuit yields contradictory results. For example, this argument is evidenced in the machine learning paper, *Inherent Trade-Offs in the Fair Determination of Risk Scores*.⁸⁰

In this paper, Kleinberg et al show that the numerous definitions of "fairness" used while discussing the pros and cons of ML algorithms are mostly incompatible with one another, except for rare specialized cases.⁸¹ This means that even if developers make a conscious attempt to fulfill a fairness condition while implementing an algorithm, they are bound to be criticized by others who hold another definition of fairness.

The three notions of fairness do not seem to contradict one another from the outset. The three properties basically correspond to the following claims. First, scores of the algorithm should have the same meaning for each data point, regardless of its group. Second, the "negative class" should not be targeted in a way that is unwarranted by the properties of the class. Third, the "positive class" should not be awarded in a way that is unwarranted by the properties of the class.⁸² Although the paper is written mainly for risk assessment systems, it is possible to translate the meaning of different classes to the field of targeted ads. Certain political affiliations, racial groups,

⁸⁰ Kleinberg, Jon, Sendhil Mullainathan, and Manish Raghavan. "Inherent trade-offs in the fair determination of risk scores." *arXiv preprint arXiv:1609.05807* (2016).

⁸¹ Ibid.

⁸² Ibid.

or socio-economical class groups can be considered the positive or negative class depending on their perception in the world of social media.

The paper continues to show that each of the three rules not only results in different scores, the scores are at odds with one another. For the cases discussed in this work, this claim displays that no algorithm could result in an optimization algorithm that satisfies all of its users. Thus, the authors demonstrate that social sciences do not yet have a common ground understanding of what fairness is, and in fact, the existing definitions yield incompatible results.

A further evidence of the claim that "better algorithms" are an unreasonable approach to social media criticism is that human decision making is as flawed as its robotic counterpart, if not more. Discussing issues of fairness in a theoretical framework might require that we compare AI decision-making to an ideal. However, from a design framework, all that we have to work with is the existing status quo, which in the case of decision making, is human beings. Therefore, one might demand that we question how reliable humans are as judges when making claims about AI judgment.

Indeed, in *Extraneous Factors in Judicial Decisions*, Danziger et al show statistically significant differences in ruling depending on whether a judge took a food break before a hearing or not.⁸³ In the paper, they extrapolate that numerous extraneous factors that should be irrelevant to the case make significant changes in courts of law, where the cost of being impartial is higher than in a social media recommendation algorithms. Similarly, the National Bureau of Economic Research shows evidence that asylum judges, loan officers and baseball umpires consistently fall into "gambler's fallacy," which is the misconception that the outcome that has been produced less

⁸³ Danziger, Shai, Jonathan Levav, and Liora Avnaim-Pesso. "Extraneous factors in judicial decisions." *Proceedings of the National Academy of Sciences* 108, no. 17 (2011): 6889-6892.

is more likely to be produced in the future, even though the trials are independent.⁸⁴ Finally, numerous studies explain that in many fields, AI decision making is more unbiased than human decision making.⁸⁵ Therefore, one might claim that the argument for flawless algorithms is unreasonable and unproductive.

Notwithstanding the legitimacy of the aforementioned claims, advancing optimization algorithms does not necessitate making the algorithms fairer or less biased. Even if making uncontested progress in optimization algorithms is difficult, manufacturing explainable algorithms is still within the capabilities of social media platforms.

In their paper titled *Accountability of AI Under the Law: The Role of Explanation*, Finale Doshi-Velez et al explain that we could reasonably expect as much transparency in decision making of algorithms as we do from humans.⁸⁶ In this work, the authors claim that it is possible to keep algorithms accountable without understanding their inner workings in total. Taking the precedents of legal perspectives as what explanations are generally considered valid in human judgment, the paper suggests that the same form of explanation can be expected of AI algorithms as well. The two methods it employs are "local explanation" and "constructing counterfactual scenarios."⁸⁷

The first insight the paper employs is that explanation and transparency are not equivalent. For example, a social media platform could expose its source code, but this would not necessarily

⁸⁴ Chen, Daniel L., Tobias J. Moskowitz, and Kelly Shue. "Decision making under the gambler's fallacy: Evidence from asylum judges, loan officers, and baseball umpires." *The Quarterly Journal of Economics* 131, no. 3 (2016): 1181-1242.

⁸⁵ "The Challenge from AI: Is 'human' Always Better?" Points. May 24, 2018. Accessed March 26, 2019. <https://points.datasociety.net/the-challenge-from-ai-is-human-always-better-f7ee0f382550>.

⁸⁶ Doshi-Velez, Finale, Mason Kortz, Ryan Budish, Chris Bavitz, Sam Gershman, David O'Brien, Stuart Schieber, James Waldo, David Weinberger, and Alexandra Wood. "Accountability of AI under the law: The role of explanation." *arXiv preprint arXiv:1711.01134* (2017).

⁸⁷ Ibid.

translate to an explanation interpretable by humans without challenging calculations. However, the authors suggest that we do not need transparency in order to get explanations for particular decisions. Instead, giving local explanation is the ability of an algorithm to respond to questions such as: "What were the main factors in a decision?" and "Why did two similar-looking cases get different decisions?".⁸⁸

A local explanation is an "explanation for a specific decision, rather than an explanation of the system's behavior overall."⁸⁹ Even though algorithms look for an abundance of parameters and give them weights through non-linear functions, the study shows that the algorithms could be questioned about particular decisions in these human-interpretable ways. Similarly, constructing counterfactual scenarios is the algorithm's ability to respond to the question: "Would changing a certain factor have changed the decision?"⁹⁰ By simulating the algorithm with an altered particular parameter, the platform could accurately estimate whether or not a particular parameter was a tipping-factor in the judgement. Through these two tools, the authors suggest that "demanding explanation from AI systems in such cases is not so onerous that we should ask of our AI systems what we ask of humans."⁹¹

It is not difficult to imagine how designing a platform with such explanations regarding its news ranking or suggestions algorithm can help with the aforementioned problems of disempowerment. Consider the videos Tufekci watches on YouTube for her research, and how she finds that she finds more extreme versions of the same category of videos in her "Up Next."⁹² With an algorithm that allows for explanations, Tufekci could ask whether a particular video would be

⁸⁸ Ibid.

⁸⁹ Ibid.

⁹⁰ Ibid.

⁹¹ Ibid.

⁹² Tufekci, *Dystopia*.

in the Up Next if she had not seen another video at a past instance. Even as it is today, YouTube allows its users to delete any video from their watch history, so that the recommendations no longer take those data points into consideration. In a way, YouTube can "forget" that the user watched the video.⁹³ In a platform that grants local explanations, Tufekci could identify the videos that have pigeonholed her into a certain category and remove them from her watch history. Similarly, Zuckerberg recently claimed that Facebook is building a system that will allow the users to delete any portion of their profiles, in a similar method to YouTube, such that the recommendation algorithms "forget" about that past behavior.⁹⁴ Users could combine local explanations with the ability to erase parts of their history to guide their own social media experiences. Thus, without having to give users physical control over the functioning of the algorithms, which might be a technical burden and might poorly affect the efficiency of the algorithms, platforms could still grant the users the ability to change the system's behavior in predictable and useful ways.

Admittedly, one drawback of exposing any information in a proprietary system is that it introduces a vulnerability into the system—namely that users are better able to manipulate the system for a desired result. For any given system, there is an underlying threat model, and in any threat model, there are actors who wish to abuse the system. The threat actors for a social media platform might be malicious users who wish to get a harmful message to a specific audience or advertisers who wish to discriminate against users in a way that is prohibited by law.

Local explanations of the sort that Doshi-Velez describe may turn out to be instance-specific enough that threat actors are not significantly bolstered by this new information.

⁹³ Lee, Joel. "Sick of Irrelevant YouTube Recommendations? Here's What You Need to Do." MakeUseOf. September 14, 2013. Accessed March 26, 2019. <https://www.makeuseof.com/tag/sick-of-irrelevant-youtube-recommendations-heres-what-you-need-to-do/>.

⁹⁴ Constine, Josh, *Zuckerberg Highlights*.

Nonetheless, it must be noted that there is an inherent danger with exposing any new information in a system, and the tradeoff requires careful deliberation from the side of developers.

The Significance of Accessibility

In another argument, one might combat the understanding that accessibility of individual features is significant in design. Instead, one might claim that the existence of a feature is sufficient for the users to engage with it. As logically advanced creatures, humans should not be confused by extra steps so much so that they are discouraged from using certain features. Especially in the case of notification settings, one might argue that creating meta-configurations that a user can curate himself is inconsequential. It should not matter whether certain decisions are opt-in or opt-out, given that there exists a way for the user to curate the settings that are desirable for him.

Behavioral economist Richard Thaler names the predictable ways in which human behavior deviates from the dictates of logic as "nudge theory."⁹⁵ The mantra of this theory has been summarized as: "If you want people to do something, make it easy."⁹⁶ Even if there is a set of decisions that are more advantageous for a person, she may fail to make this choice and settle with the default—given that there are barriers to changing behavior. For example, in his study outlining the measures necessary to increase enrollment to retirement plans in the US, Thaler claims: "There is now conclusive evidence that automatic enrollment, where employees are automatically signed up unless they opt out, is extremely successful in overcoming the procrastination that can impede signing up."⁹⁷ The study eliminates other possible explanations to make the claim that the increase is decisively tied to the default *opt-out* strategy. One can see that

⁹⁵ "If You Want People to Do Something, Make It Easy." Open Box Technology. October 03, 2017. Accessed March 26, 2019. <http://openboxtechnology.com/if-you-want-people-to-do-something-make-it-easy/>.

⁹⁶ Ibid.

⁹⁷ Benartzi, S., & Thaler, R. (2013). Economics. Behavioral economics and the retirement savings crisis. *Science (New York, N.Y.)*, 339(6124), 1152-3.

if individuals are largely unwilling to make an active effort to change the default behavior in a long lasting and impactful decision such as retirement savings, it is difficult to imagine that users would be more willing to make an active effort to change their notification settings—an admittedly smaller-scale and less impactful decision.

Furthermore, it is not only opting decisions that change ultimate results, but any other obstruction as well. Thaler's paper outlines other results that reaffirm the potential impact of this suggestions, one of which is as follows: If forced to actively decide rather than passively accept the default investment rates, employees prefer higher investment rates, which is assumed in the paper to be a desired result.⁹⁸ Similarly, it is safe to assume that if users are taken to the notification settings page when they first sign up, they will be more likely to curate settings that are to their benefit, whatever they might be in individual cases.

The Tradeoff between Flexibility and User-Friendliness

A final argument may claim that increasing options is far from an ideal solution, since it introduces new problems of user-friendliness. Indeed, if there was no such tradeoff, any technology could introduce more granular preferences ad infinitum. For notification settings, as well as for other areas of social media design, there is a clear downside to adding complexity to decision making.

Hick's law suggests that the increased number of options in a given environment corresponds to an increased time that the user takes to make a decision.⁹⁹ Giving more options has the tradeoff of requiring the actor to consider a larger field of possibilities, increasing the likelihood of overwhelming the user. This becomes a problem in the case of software design especially

⁹⁸ Ibid.

⁹⁹ The Interaction Design Foundation, *The Pareto Principle*.

combined with another fact: Human attention is a scarce and fragile resource that is easily distracted, especially when challenged.¹⁰⁰ Therefore, Hick's law is a rule of thumb in human computer interaction which anticipates that engagement and conversion rates tend to drop when a technology introduces numerous options without consideration for user friendliness.

Particularly in the case of notification settings, the user could easily fall on the default settings in order not to make decisions that challenge his attention. One could argue that the increased number of options only creates a perception of choice, whereas most users will most likely continue with the default settings. As the Interaction Design Foundation puts it: "Users bombarded with choices have to take time to interpret and decide, giving them work they don't want."¹⁰¹ Furthermore, if decision making is required for the user to continue with the website, the effect of the bombardment can go as far as pushing the user to close the website to delay the decision making.

There is no denying the Hick's law is a fundamental design principle that must be considered. Design requires finding the right balance between tradeoffs; compromises must be made to ensure a convenient medium between complexity and simplicity. Thus, other resources of design theory hints at the possibility of mitigating the added complexity of more choices.

The Interaction Design Foundation explains that if more options are essential, grouping choices by high-level categories and progressively disclosing more details allows the user to be familiar with the field of possibilities in every screen. If the user knows what they want before encountering the screen, having more choices does not have the same effect. Therefore, it is useful

¹⁰⁰ Horvitz, Eric, Carl Kadie, Tim Paek, and David Hovel. "Models of Attention in Human-Computer Interaction: From Principles to Applications." *interaction* 3, no. 4: 18-19.

¹⁰¹ The Interaction Design Foundation, *The Pareto Principle*.

to break down the decision processes such that going into each screen, the user already knows what choice he wishes to make.¹⁰² Hence, the notification settings recommended in the last chapter provide a small number of meta-configurations, after which the user can make more detailed decisions if he chooses to. By grouping all notification settings in three broad categories, the user may enjoy the functionality without being overwhelmed by the added choices.

Furthermore, by grouping the items in these specific categories, the recommendation makes use of an aforementioned design principle called the "Pareto Principle" or the "80/20 Rule." Although the rule has been applied to all types of data, within human computer interaction, it argues that 80 percent of users use merely 20 percent of the functionality.¹⁰³ In the case of notification settings, the recommendation extrapolates that the needs of most users are covered by the two meta-configurations, "permissive" and "conservative". The rest of the functionality is cloaked by the "power user" configuration, which is aimed at the remaining 20 percent of the users.

Having responded to the anticipated counter-arguments, the rest of the work provides a summary of the farther-reaching implications of the claims made thus far.

¹⁰² Ibid.

¹⁰³ Ibid.

Conclusion

Throughout history, mediums of mass communication have been held up to scrutiny, be it the newspaper, the radio, the television, or the "global village."¹⁰⁴ It is famously known that Socrates was against writing, because he believed the medium would destroy authentic dialogue, threaten the intimacy and soul-to-soul nature of communication, and provide "not truth, but only the semblance of truth."¹⁰⁵ Nonetheless, tools of communication have continued to develop, and each medium has been offered a unique critique that stemmed from its situatedness in a particular time, space and specific circumstances.

Today, social media is uncontestably one of the most popular forms of mass communication. As expected, many thinkers challenge every aspect of the medium and find no redeeming qualities in it. In the field of political science, Vaidhyathan claims that the problem with Facebook "is Facebook" itself, and he argues that the only way to fix the irreparable damages Facebook has caused is to close down the platform entirely.¹⁰⁶ In philosophy, Dreyfus suggests that the nihilism Internet has harbored can only be avoided if one turns off the monitor and finds a hobby or cause to guide her human endeavors.¹⁰⁷ In sociology, Tufekci likens the ad-based platform to an AI-powered dystopia.¹⁰⁸ These thinkers are not especially radical; to the contrary, they are the authors of some of the most well-articulated and sound critique social media faces today.

However, the techno-pessimism that is easy to slip into is not only an unproductive approach towards improving the medium, but it is also misdirected. The concept of social media does not require that its content be categorized by posts of a certain format, distributed in a specific

¹⁰⁴ Dreyfus, *OI*.

¹⁰⁵ Plato, C. J. "Phaedrus." (1988).

¹⁰⁶ Vaidhyathan, *AM*.

¹⁰⁷ Dreyfus, *OI*.

¹⁰⁸ Tufekci, *Dystopia*.

algorithm, or accessed through a particular interface. These are decisions that creators of today's platforms have made—decisions that are continually made.

At its genesis Facebook did not have a timeline, and its home page only displayed activity the user was involved in. Instagram only established "stories" in 2016, after a completely separate platform—Snapchat—was built around the idea. Twitter famously started with a 140-character limit, which was doubled as it became evident that users built ad hoc solutions to get over the limit, such as numbering their posts so that they could be read as parts of a single piece. Some of these changes may seem trivial at first glance, but the crucial point is that social media platforms are not constants we have to either accept or reject from the outset. Rather, they are subject to change and influence through incentives.

Therefore, the ambition of this thesis to redirect the condemnation serves a two-fold goal; one that is redemptive, and one that is prescriptive. The redemptive goal is to save the medium from the criticism that taints the platforms built within it, because the wholesale rejection of any system disallows progress and hinders advancement, replacing these with the singular aim of the disposal of the system. Uncritical acceptance of techno-pessimism, as well as techno-optimism, pushes its followers to agitated and radical solutions, be it Luddism or Technologism. Of course, this charge is not directed at the critics consulted in this work. Indeed, it has been one of the tasks of this work to prove and further advocate the validity of their comments. However, the grand scale critique of the first-generation of thinkers of social media must give way to a second generation—one that improves on their methods of evaluation in the direction of more specificity, more nuance and more individuation. Thus, the primary goal of this thesis is to save social media from criticism that is better understood when pointed at platforms.

Indeed, the bifurcation of the medium and its realizations is not a mere formalism. The second goal of this work, which is prescriptive, is to show that the content of the platform can and should be updated in the directions provided. Social media platforms awaken new anxieties in humans and confuse us as to when we are alone and when we are in a community. As phones continue to take up more space in our lives, the main gate between the physical and the online world, the notification feature, increases in significance. As nudge theory suggests, what is easily accessible is decisive in what users engage with. The lack of a rich set of options as to when and how users can engage with their social media platforms has real-life consequences that are far from ideal. Even after the users enter the platform, the magnificent world inside brings forth its own discontents. The structure in which content is presented falls contrary to how humans make the world intelligible to themselves. In return, users encounter newly found vices of the online world, ranging from fake news to a conflation of ads with genuine content, and from the unintended levelling of different types of content to a non-personalized objectification on a massive scale. Finally, beyond the interface that is visible to the user, the fashion in which content is curated similarly brings an increased incentive to sensationalize and radicalize one's thoughts. It causes an impairing of exploratory attitudes, replacing them instead by narrow and confined echo chambers.

Sheltering the medium from this criticism is not intended to encourage a policy of inaction towards social media. On the contrary, it is meant to find productive paths towards the improvement of platforms without giving up on the overarching ideal of global communication. Therefore, our recommendations are concrete, even if they require giving up on grandiose statements, and instead focusing on the system's minute details. Notification settings can incorporate a richer set of options, while remaining conscious of the trade-off between flexibility and usability. Deceptive ads, fake news, and the experienced need to appeal to the least common

denominator can be countered by constructing separate communities within platforms and allowing independent communal norms, as well as by varying the design of the publishing of content. Finally, designing appropriate optimization functions can retaliate the radicalization and pigeonholing of the content the user encounters and generates.

Each claim made in this work is subject to criticism. Neither the mapping of the critique to design features, nor the proposed alternative architectures is perfect. Nonetheless, its redemptive and prescriptive functions remain as a way forward, albeit as a set of flawed blueprints rather than a full-fledged roadmap. Thus, this work is presented as a proof of concept; namely for the claim that the grand-scale criticism of social media can lend itself to betterment, given that we direct our attention to specific and minute design choices, rather than the intrinsic limitations of an entire medium. Though it requires more effort from thinkers, users and platforms alike, we must continuously imagine new ways of architecting these online worlds so that we are not despaired by the self-crippling notion that an entire method of communication is detrimental to the fabric of society. Instead, through humble but persistent revisions of social media that have been outlined in this work, we can create a healthier tool for public communication, and pave the way towards the redemption of a now-trivialized and contaminated ideal of global interconnectedness.

Bibliography

- Babulous. "Why I Kicked Facebook off My Phone." Hacker Noon. April 15, 2017. Accessed March 26, 2019. <https://hackernoon.com/why-i-kicked-facebook-off-my-phone-a254c5b410bc>.
- Benartzi, S., & Thaler, R. (2013). Economics. Behavioral economics and the retirement savings crisis. *Science (New York, N.Y.)*, 339(6124), 1152-3.
- Bleich, Sara N., Julia A. Wolfson, and Marian P. Jarlenski. "Calorie changes in chain restaurant menu items: implications for obesity and evaluations of menu labeling." *American journal of preventive medicine* 48, no. 1 (2015): 70-75.
- Carman, Tim. "Fast Casual Nation: The Movement That Has Changed How America Eats." The Washington Post. August 29, 2017. Accessed March 26, 2019. https://www.washingtonpost.com/lifestyle/food/fast-casual-nation-the-movement-thats-changing-how-we-eat/2017/08/28/23f6d710-86c5-11e7-961d-2f373b3977ee_story.html?noredirect=on&utm_term=.289b1ab110f1.
- "The Challenge from AI: Is "human" Always Better?" Points. May 24, 2018. Accessed March 26, 2019. <https://points.datasociety.net/the-challenge-from-ai-is-human-always-better-f7ee0f382550>.
- Chen, Daniel L., Tobias J. Moskowitz, and Kelly Shue. "Decision making under the gambler's fallacy: Evidence from asylum judges, loan officers, and baseball umpires." *The Quarterly Journal of Economics* 131, no. 3 (2016): 1181-1242.
- Constine, Josh. "Highlights & Transcript from Zuckerberg's 20K-word Ethics Talk." TechCrunch. February 20, 2019. Accessed March 26, 2019. <https://techcrunch.com/2019/02/20/zuckerberg-harvard-zittrain/>.

"Continuing Our Work to Improve Recommendations on YouTube." Official YouTube Blog.

January 25, 2019. Accessed March 26, 2019.

<https://youtube.googleblog.com/2019/01/continuing-our-work-to-improve.html>.

Danziger, Shai, Jonathan Levav, and Liora Avnaim-Pesso. "Extraneous factors in judicial decisions." *Proceedings of the National Academy of Sciences* 108, no. 17 (2011): 6889-6892.

Davis, Hank, and S. Lyndsay McLeod. "Why humans value sensational news: An evolutionary perspective." *Evolution and Human Behavior* 24, no. 3 (2003): 208-216.

Doshi-Velez, Finale, Mason Kortz, Ryan Budish, Chris Bavitz, Sam Gershman, David O'Brien, Stuart Schieber, James Waldo, David Weinberger, and Alexandra Wood. "Accountability of AI under the law: The role of explanation." *arXiv preprint arXiv:1711.01134* (2017).

Dreyfus, Hubert L. *On the internet*. Routledge, 2013.

Facebook. Accessed March 26, 2019.

https://www.facebook.com/settings?tab=notifications&ion=on_facebook&view.

Gershon, Ilana. "Un-friend my heart: Facebook, promiscuity, and heartbreak in a neoliberal age." *Anthropological Quarterly* 84, no. 4 (2011): 865-894.

Habermas, Jurgen, Jürgen Habermas, and Thomas Mccarthy. *The structural transformation of the public sphere: An inquiry into a category of bourgeois society*. MIT press, 1991.

Heidegger, Martin. "The Thing." *Poetry, Language, Thought.*, Harper & Row, 2001.

Hendler, James, and Jennifer Golbeck. "Metcalf's law, Web 2.0, and the Semantic Web." *Web Semantics: Science, Services and Agents on the World Wide Web* 6, no. 1 (2008): 14-20.

Horvitz, Eric, Carl Kadie, Tim Paek, and David Hovel. "Models of Attention in Human-Computer Interaction: From Principles to Applications." *interaction* 3, no. 4: 18-19.

"If You Want People to Do Something, Make It Easy." Open Box Technology. October 03, 2017.

Accessed March 26, 2019. <http://openboxtechnology.com/if-you-want-people-to-do-something-make-it-easy/>.

"Introducing "You're All Caught Up" in Feed." Instagram. July 02, 2018. Accessed March 26, 2019. <https://instagram-press.com/blog/2018/07/02/introducing-youre-all-caught-up-in-feed/>.

Kleinberg, Jon, Sendhil Mullainathan, and Manish Raghavan. "Inherent trade-offs in the fair determination of risk scores." *arXiv preprint arXiv:1609.05807* (2016).

Lapowsky, Issie. "The 21 (and Counting) Biggest Facebook Scandals of 2018." Wired. December 21, 2018. Accessed March 26, 2019. <https://www.wired.com/story/facebook-scandals-2018/>.

Lee, Joel. "Sick of Irrelevant YouTube Recommendations? Here's What You Need to Do." MakeUseOf. September 14, 2013. Accessed March 26, 2019. <https://www.makeuseof.com/tag/sick-of-irrelevant-youtube-recommendations-heres-what-you-need-to-do/>.

"Login Button - Facebook Login - Documentation." Facebook for Developers. Accessed March 26, 2019. <https://developers.facebook.com/docs/facebook-login/web/login-button/>.

Martonik, Andrew, and Andrew Martonik. "How to Turn off the New Persistent Facebook Notification." Android Central. May 10, 2013. Accessed March 26, 2019. <https://www.androidcentral.com/how-turn-new-persistent-facebook-notification>.

McAndrew, Frank T. "Can gossip be good?." *Scientific American Mind* 19, no. 5 (2008): 26-33.

McLuhan, Marshall, and Quentin Fiore. *The Medium Is the Massage: an Inventory of Effects*. Penguin, 2008.

McLuhan, Marshall. *Understanding Media; the Extension of Man*. McGraw-Hill Book Company, 1965.

Merrill, Jeremy B. "What We Learned from Collecting 100,000 Targeted Facebook Ads."

ProPublica. March 09, 2019. Accessed March 26, 2019.

<https://www.propublica.org/article/facebook-political-ad-collector-targeted-ads-what-we-learned>.

Nickinson, Phil. "Facebook Quietly Adds Facebook Home Support (and Unofficial Support) for Galaxy S4, HTC One, Others." Android Central. May 10, 2013. Accessed March 26, 2019. <https://www.androidcentral.com/facebook-quietly-turns-support-and-unofficial-support-galaxy-s4-htc-one-others>.

"Notification Basics & Settings | Facebook Help Center." Facebook. Accessed March 26, 2019.

<https://www.facebook.com/help/327994277286267>.

"The Pareto Principle and Your User Experience Work." The Interaction Design Foundation.

Accessed March 26, 2019. <http://www.interaction-design.org/literature/article/the-pareto-principle-and-your-user-experience-work>.

Plato, C. J. "Phaedrus." (1988).

Postman, Neil. *Technopoly: The Surrender of Culture to Technology*. Vintage Books, 1993.

Przybylski, Andrew K., et al. "Motivational, emotional, and behavioral correlates of fear of missing out." *Computers in Human Behavior* 29.4 (2013): 1841-1848.

Rodriguez, Salvador. "Here Are the Scandals and Other Incidents That Have Sent Facebook's Share Price Tanking in 2018." CNBC. November 21, 2018. Accessed March 26, 2019.

<https://www.cnn.com/2018/11/20/facebooks-scandals-in-2018-effect-on-stock.html>.

Sartre, Jean-Paul. "The Look". *Being and nothingness*. Open Road Media, 2012.

Tufekci, Zeynep. "Machine intelligence makes human morals more important." *TED Talk* (2016).

Tufekci, Zeynep. "We're building a dystopia just to make people click on ads." *TED: Ideas Worth Spreading*, Sept (2017).

Tufekci, Zeynep. "Yes, Big Platforms Could Change Their Business Models." *Wired*. December 17, 2018. Accessed March 26, 2019. <https://www.wired.com/story/big-platforms-could-change-business-models/>.

Turkle, Sherry. *Alone Together: Why We Expect More from Technology and Less from Each Other*. Basic Books, 2017. 329.

V Shred's Facebook Page, Accessed March 26, 2019.

<https://www.facebook.com/watch/?v=2267436653579031>

Vaidhyanathan, Siva. *Antisocial media: How Facebook disconnects us and undermines democracy*. Oxford University Press, 2018.

"What Is a Filter Bubble? - Definition from Techopedia." *Techopedia.com*. Accessed March 26, 2019. <https://www.techopedia.com/definition/28556/filter-bubble>.

Wineburg, Sam, Sarah McGrew, Joel Breakstone, and Teresa Ortega. "Evaluating information: The cornerstone of civic online reasoning." *Stanford Digital Repository*. Retrieved January 8 (2016): 2018.

Wong, Julia Carrie. "How Facebook and YouTube Help Spread Anti-vaxxer Propaganda." *The Guardian*. February 01, 2019. Accessed March 26, 2019. <http://www.theguardian.com/media/2019/feb/01/facebook-youtube-anti-vaccination-misinformation-social-media>.

Zittrain, Jonathan. *The Future of the Internet--and How to Stop It*. Yale University Press, 2009.