**w**hat **w**ent **w**rong: 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

**Chapter 1: Introduction**

The history of the perceptions of new media of communication is variations on the same phenomena: The early enthusiasts who make a grand claim about the new world of possibilities the new medium has opened up, followed by the critics who claim that the world the new medium has created is an arm's reach from a dystopia. When the newspaper appeared in the 18th century, Enlightenment thinkers praised it as the necessary grounds for a public sphere in which individuals can engage in rational reflection and advance as their societies.[[1]](#footnote-1)Soon after, Kierkegaard castigated the very same medium for creating a place for pure reflection with no place for action.[[2]](#footnote-2) Following the global acclaim of the television that would finally bring voice to a global audience, Heidegger pessimistically noted on the impeding loss of any understanding of closeness of ideas or people.[[3]](#footnote-3)

Similarly, the early enthusiasts of the Internet imagined that the ease of communication it would eventually allow would bring human beings closer to one another, allow for more meaningful interactions between those who had previously 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 romantic "cyberspace" and unbounded "information highway" were used to articulate the beginning of a wholesome interconnectedness that was rivaled by no other age.[[4]](#footnote-4)

Following its historical predecessors, the Internet soon met its critiques as well; the tecno-pessimists. Among the first arguments against the Internet was that it would make all alternative forms of communication irrelevant and declare its reign as a "technopoly".[[5]](#footnote-5) The argument made its way to the question of whether or not *Second Life,* a role-playing computer game, would take the place of the physical world.[[6]](#footnote-6) Today, the techno-pessimist critique targets social media most fiercely. The prevalent claims against social media include that it is not a platform of mass communication, but mass dissemination,[[7]](#footnote-7) that YouTube is creating a dystopia for the sake of higher ad revenue,[[8]](#footnote-8) and that Facebook has undermined democracy in an irreparable way[[9]](#footnote-9). In the fields of social anthropology, tech sociology and political science, social media takes the fault for a considerable part of the prominent polemic of our day.

There is no doubt that the critique is insightful. Theoretical discussions are supported by empirical epidemics. In 2018 alone, Facebook has received public outcry for allowing foreign influence to enter 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]](#footnote-10) Although Facebook seems to be the most scandal-ridden of the platforms, each platform receives its share of public outcries. Even during the first months of 2019, a new scandal has emerged as YouTube is partially blamed for the rise of anti-vaccination movements, one of the top 10 global health threats according to the World Health Organization.[[11]](#footnote-11) Both supported by theory and in practice, there are clear problems with social media that need to be addressed.

However, the critique also often misses a critical point of distinction, namely that between what is inherent in the medium and what is added by design. In other words, the critique of social media, while insightful, is misdirected at the medium itself, whereas it should target particular design decisions, which can be fixed with alternate design forms. As a set of applications that live on top of the World Wide Web, which itself lives as an application on top of the Internet, social media is only bound by the rules of the rules of communication of the Internet. Furthermore, 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]](#footnote-12) In other words, there is a wide freedom left for the applications on the Internet to fill through their own design. As a conclusion, in how social media platforms behave, little is due to the nature of the medium, while much is due to design choices made by the particular platforms.

Therefore, the question of social media is not a medium question, but a design question. What is meant by design not a set of aesthetic decisions, as the word is colloquially used, but a set of structural and architectural decisions, in the technical sense of the word. Anything that forms the frame around the content of the platform is considered design. As Marshall McLuhan writes in a famous piece on media theory: "The medium is the message".[[13]](#footnote-13) Any content within a medium cannot be separated from the structure of the medium. Similarly, for the case of social media, any interaction a user has with a social media platform cannot be separated from the design of the platform. Whether or not it is intentional, the design of the platform already nudges the user towards certain behaviors, encourages some interactions over others, and eventually has an undeniable influence over the totality of ways in which users use the platform.

  Furthermore, design not only refers to what is made possible and what is not; but also, to what is made easily accessible and what is not. Perhaps the theoretically conscious and technically skilled user can make use of the full spectrum of the platform's possibilities, but most users will limit themselves to the uses that they see as clear and beneficial. The "80/20 rule" of human-computer interaction suggests that 80 percent of the users will use 20 percent of the functionality of a platform, which make up the easily accessible and basic features of the platform.[[14]](#footnote-14) What the platform holds is not only a question of what is available to the user, but also of what is easily accessible. Therefore, the evaluation of the critique of social media and its response should be approached as a design problem.

In the following pages, we discuss some of the central critique of social media, some of which has been alluded to already. We attempt to rescue the medium itself from the critique, instead directing it at particular design decisions that we argue have given rise to the problems articulated. Admittedly, that the medium is worth saving only follows from our implicit conviction that communication is a valuable ideal that deserves advancement. Beyond its particular implementations, "social media" stands for the more general and flexible concept of globalized public communication through the Internet. We leave the discussion of the legitimacy of the ideal of global communication to works of moral philosophy. Instead, we focus on the claim that if we choose to, we have the ability to shelter this ideal of global public communication from the critique that its current realizations should face.

To assure the validity of this claim, we argue for the possibility of alternative design choices that would alleviate the critics' concerns. Specifically, we categorize a wide set of arguments into three groups, according to the design features from which they originate, namely: notification settings, structure of posted content, and engagement-based optimization algorithms. For each of these categories, we provide the said critique and give evidence to it in order to display its legitimacy. Then, we demonstrate their connection to the design feature, and give suggestions for alternative design features. In Chapter 3, we anticipate counter arguments to our claims. Once again, we give grounds for these counter arguments to display their legitimacy and provide responds to them. We categorize these counter arguments into three groups, namely: system-wide arguments, general arguments that pertain to multiple claims, and specific arguments that pertain to a singular claim. Finally, in Chapter 4, we turn to a discussion of wider implications of our claims.

**Chapter 2**

"Societies have always been shaped more by the nature of the media by which men communicate than by the content of the communication".[[15]](#footnote-15) In this chapter, we will consider some of the criticism tech sociologists have made about social media and discuss the roots of the criticism as side effects of particular design decisions rather than something inherent to online mass communication.

Sherry Turkle, one of the leading tech sociologists of today, explains the thesis of her book *Alone Together* as this: "We are confused about when we are alone and when we are together".[[16]](#footnote-16) This is a more nuanced and defendable claim than the frequent lamentation that people are on their phones all the time these days. Turkle's claim is not one about how the most recent technological device no longer allows us to communicate in more personal ways. Actually, it is not about the devices at all. Turkle suggests that while placing networked communication into our pre-networked lives, we have created virtual selves for whom some constraints of the real selves do not apply; it has certain entitlements. For example, she says of the virtual self: "It can absent itself from its physical surround [...], it can experience the physical and virtual in near simultaneity. And it is able to make more time by multitasking".[[17]](#footnote-17) In itself, these entitlements are advantages of the system, and each marks an ability that is denied to a physical self. However, the problem stems from the confusion of the user between the two selves.

Turkle hammers this point home when she agrees that "we have always found ways to escape from ourselves, neither the desire nor the possibility is new with the Internet".[[18]](#footnote-18) However, she claims, the difference that makes the Internet a worthwhile case study to investigate is its capacity to weave together the virtual and the physical, allow a simultaneous access to both; what she calls a "life mix" that is co-constituted by one's life offline and online. Clearly, this capability does not inherently harm its users, but Turkle's findings suggest that users now allow themselves to partition their attention to the fields in a sense of "continual copresence".[[19]](#footnote-19) In turn, the misuse of these technologies 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". Similarly, when misused, 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]](#footnote-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; "FoMo".[[21]](#footnote-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]](#footnote-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 capacities that are native to the system versus the misdoings in its perception are clear in Turkle's analysis. Indeed, this is what pushes us to search for a reformation in design, since design stands between the native system and the user that populates, shapes and ultimately realizes the system. The first set of design choices that are relevant in the user's misuse into a "life mix" is ones regarding notification settings.

If the main gate that connects the physical and virtual worlds is the phone screen, especially in the cases Turkle is concerned with where one "moves into the virtual [world] with fluidity and on the go", then the notification icon is the knock on the gate. In Facebook lingo, it is called the "Red Alert Notification".[[23]](#footnote-23) The Red Alert Notification serves as an indicator to the user that some activity that is relevant to her, presumably even personally addressed to her, just occurred in real time. The alert is likely to create a sense of urgency, especially because it arrives in real time and can trigger the aforementioned fear of missing out. Similarly, it is likely to grab the user's attention as a bright red bubble that places itself in the app icon and makes itself readily visible in the user's home screen. Of course, once again, the notification is a capability to alert the user to some relevant activity. However, one must take caution that while serving this purpose, the alert does not induce an anxiety in the user to enter the platform. In other words, the user must have some say in what notifications she gets. These are called notification settings.

When we investigate notification settings of the most popular social media platforms, we encounter many points of deficiency. Facebook's notification algorithm is a prime example.[[24]](#footnote-24) Settings regarding app notifications in smart phones are shared between the smart phone settings and the app settings, but phone settings are out of the scope of this thesis. In undetailed terms, smart phones generally have on/off settings for notifications of any specific app, but they do not have any customization ability. Sensibly, more detailed configurations are left to the app settings. However, looking at Facebook's notification settings, we can see that the design lacks many settings that could prevent a user from being confused in the "life-mix" between the virtual and physical worlds.

The main statement of Facebook's notification settings reads: "You can't turn off notifications entirely, but you can choose what you're notified about and how you're notified".[[25]](#footnote-25) This is a reasonable approach, given that users expect to be notified about certain activity, since we have established the ability to be connected in real time as an advantage of networked communication. However, the choice on what/how one will be notified proves to be somewhat of a burden. Using Facebook on the web, the following is the method of altering notifications: "You'll see every notification on Facebook, but you can turn off notifications about specific posts as you view them".[[26]](#footnote-26) This means that Facebook is generally opt-out; the user will get a greedy notification algorithm by default and can change it as he goes through the platform, one by one. Like any opt-out policy, to change the user must have some level of awareness and capability. Furthermore, the requirement that some settings can only be changed individually makes certain changes virtually impossible if the user has become aware of them after some time of using the app. For example, notifications from third party app requests and activity can only be turned off app by app. Since using the "Continue with Facebook"[[27]](#footnote-27) button on any third-party platform generally puts their app on the list, numbers for which this may require a significant amount of work for users.

However, even the discerning user might fail to keep up with notification settings due to Facebook's frequent app updates. New settings are introduced, or some settings are reorganized with updates, and Facebook generally defaults all of these settings to opt-out. As Android Central notes: "Facebook likes to do a lot of "opt out" rather than "opt in" changes on its app updates".[[28]](#footnote-28) This forces the discerning user to not only engage in the process of changing notification settings once, but to do it regularly, i.e., every time Facebook updates its app. This is especially a burden given that some of the updates can reasonably be perceived as intrusive: "[T]he Facebook app update has added a new notification -- a persistent "ongoing" notification that lives in your notification pull-down".[[29]](#footnote-29) A notification that does not signify any particular activity and re-appears every time it has been responded to fails to capture the purpose of a notification that we have outlined above. Although this is an extreme example that might have been a bug, the point stands that Facebook frequently updates its app, and often does not document all the changes in the description.[[30]](#footnote-30) Estimated around once to twice a week, Facebook updates its app with the same description for every update that does not provide any details into the specific update. It 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]](#footnote-31)

This is not standard practice, since other apps, even WhatsApp and Messenger that are owned by Facebook Inc detail out the specifics of each update in their update descriptions. In the end, the burden to regularly organize one's settings is left to the user, which is one of the culprits for the "life mix" that Turkle criticizes. Especially with "push notifications", which are the red alert notifications that appear in the user's device screen when she is not "actively using Facebook", the only configuration is a complete on/off that the user can change from the smart phone settings.[[32]](#footnote-32) Facebook does not have any native configurations to pause all, or any, notifications, either indefinitely or for a time being.

Therefore, opt-out notification settings may push the users to live in the "life-mix" even if the users are not aware of it. We have established that often, our journey in a social media platform begins with our intention to see what the notification is signaling. This is because the icon is not purely informative, but also has an inviting function. Once the user clicks the notification to view it, although he may view the content of the notification in isolation, it is not difficult to get sucked in the platform without noticing. This is not only true for the virtual world, but the physical world as well. Humans are wired to view the world in circumspection. As an example, we can think of a store that has a "50% discount on selected items" sign on the door. Inside the store, the items on sale are likely proximate to the rest of the items. Any shopper knows that once he enters the store, there is a good chance that his attention will move beyond the items on sale, to be grabbed by an item from new season. Once the customer's initial motivation to check out the sales is overcome, the customer is just like any other who has come to check out the whole store.

When this happens during the user's engagement with the physical world, we *experience* ourselves in the platform instead of in the outside world. Just like when we are watching a movie or reading a book, our experience is not purely determined by where we are located, but where our attention is directed. For example, we can imagine a pedestrian who is checking her phone while walking who does not seem to notice our presence. Therefore, the awareness of the physical environment is heavily reduced when one is engaged with the online world.

In an alternative platform with a different design architecture, the platform app can allow the user to make more specific changes in push notifications. To ensure the user still has access to activity that is worth interrupting her offline world without creating a sense of "continuous copresence", the app can at the very least have a "conservative" configuration that is all opt-in, meaning that its default state is no notifications. Since the default settings of current apps are all-permissive, meaning that they require the user to actively opt-out, the conservative configuration can act as a balance and be all opt-in.

Furthermore, this allows users that are less willing/able to change specific configurations to turn off all notifications at times without dealing with the smart phone settings separately, and users that are more willing/able can configure their conservative settings to their liking. All the meanwhile, users that do not wish to engage in any notification restrictions can have a similar experience to the current state without being bothered by these more advanced settings. The tradeoff between opt-in and opt-out is varied, and its complications are discussed in Chapter 3 in the form of a counter argument. In short, the meta-options of "permissive", "conservative" and "power user" can be used to create a balance between flexibility and ease of use. Thus, the user is granted capability to remain offline for a period of time and has a choice in how connected he will be during general use. Finally, the user can be taken to a settings page at the time of signing up to be informed on the default opt-out policies, as this is a natural place for initial configuration settings. She can be introduced to the three options aforementioned during the sign up process as well.

A more radical type of techno-pessimism is found in Siva Vaidhyanathan's *Antisocial Media,* where he seems to argue that "the problem with Facebook is Facebook".[[33]](#footnote-33) However, when one examines his arguments, which are valid and well-put, it is clear that there is nothing inherently problematic in social media, and his polemic is with Facebook's design. Of course, we can generalize Vaidhyanathan's arguments to more social media platforms than just Facebook.

Vaidhyanathan recognizes that in Facebook, there is a difficulty to distinguish between different types of content.[[34]](#footnote-34) 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 event type, a text, an optional visual component (a photo or a video), and the engagement bar (which includes the options: like, share, comment). This invariant structure that is encouraged also by the main form of engagement that is "scrolling through"[[35]](#footnote-35) the feed creates a confusing whirlwind of posts for the spectator. This fact generalizes to Instagram, Snapchat and Twitter 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 more. However, besides small differences, the invariant structure is the same in all of these platforms.

It is a fact of educational psychology that we find it easier to make sense of information when we also have a context for that information, as opposed to a piece of information that is given out of context. When discontinuous posts with similar formats but widely ranging content becomes the units through which we experience social media, the possibility of a context, or a gradient of a story becomes impossible. If it is a goal to retain the information we received during a visit to social media or have an overall understanding of our experience that we keep after the visit, we would need to group relevant posts together. This thesis will not present evidence that users of Facebook have this kind of a goal. Indeed, perhaps users do not want to seriously engage with the platform, but to use it as a distraction without much regard for a longer-term consequence. However, given the claim that social media's potential exceeds its current main use of pure entertainment and distraction, the grouping of the posts (or lack thereof) and the resulting confusion of the user throughout a visit to the platform is a possible culprit.

Furthermore, sponsored ads closely mimic regular posts in the feed. In Facebook, sponsored content has the same three sections, with the only difference being a "sponsored" sub-text. To blend in with regular posts, ads often use videos or news articles that do not make it clear from the initial engagement that they are sponsored content. For example, a body-building app ad presents itself as a regular engagement page with a video and a survey for its users, which reveals itself to be an ad at the end of the survey.[[36]](#footnote-36) Similarly, Instagram's ads are photos posted in the same style, except with a clickable caption, and Snapchat stories include sponsored videos that often begin as a regular story and then reveal themselves to be ads.

This is a useful marketing strategy, since a good number of people have a negative gut reaction to ads, and therefore are more likely to retain information if they first do not recognize it as sponsored. Especially when ads are targeted, the content is often relevant to the users' interests. Therefore, it is not surprising that one might not be able to discern between sponsored content and regular content during a quick scroll-through. Indeed, a recent Stanford University study observed that "more than 80 percent [of participating students] believed a native ad, identified with the words 'sponsored content,' was a real news story".[[37]](#footnote-37) Along with the confusion induced in the user due to the invariant structure of the posts, the indistinguishability of the sponsored content from regular content is prone to creating a distrust in the user, since she may have the experience of engaging with a post that she believes is "genuine", only to find out that it was an ad. Furthermore, 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".[[38]](#footnote-38) Once again, this is a sensible strategy for advertisers, and it makes sense for social media platforms to take advantage of it. However, Vaidhyanathan's claim that users being unable to differentiate between ads and native content is in some way detrimental is insightful and deserves design-oriented investigation.

The invariant structure affects the way social media is used both by the posters and the "lurkers" (a word Internet culture uses to mean spectator). Since the less and more significant types of content are identical in structure, it encourages the lurkers to pay little attention to any given post. Given that the pure entertainment, personally intimate and politically significant are mashed together in a complete similarity that requires active attention to separate, the lurker ends up glancing through the feed, only stopping for the posts that catch her eye. Most of the time, the posts that catch her eye are the more sensationalists ones -- the posts that use a flashy picture or a radical word. However, the attention granted even to these posts cannot be too long, because due to the scrolling-feed, the next post already makes itself visible while the user takes a look at the current post. Furthermore, due to the practically unending home feeds, the user has virtually infinite content that she can scroll through. Therefore, even if she chooses to engage, she is likely to click one of the reaction buttons (like, share, other emoticons), or write a short comment. In the end, leaving the platform, it is difficult for the user to come away with having retained any of the information she has observed. As Vaidhyanathan puts it a bit harshly: "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]](#footnote-39)

The invariant structure has a similar trivializing effect on how the poster's use the platform. Since posters are lurkers are not separate groups but the same individuals at different times, the poster is acutely aware of how lurkers scroll through the feed. Therefore, if he wants to be read and engaged with, which we can reasonably assume a poster of a social media platform does, he has to create one of the posts that catch the lurkers' eyes. As previously mentioned, these are the more sensationalist posts that include a portion that is outrageous, unfamiliar, or in some way extreme, so that it jumps out of the rest of the feed with innumerable posts. The lurkers' behavior encourages the posters' behavior, and vice versa, so that in the end, a vicious cycle of lack of attention and reflection from users is sustained.

Furthermore, another criticism targeted at social media falls under the same design choice. In *On the Internet,* Hubert Dreyfus warns us against the levelling effect of organizing a platform without a concern for inter-post context.[[40]](#footnote-40) 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. During a scroll through, it takes conscious work for a user to identify what group any post belongs to, so she must view them on the same grounds. Then, since the post that is not significant (say, the puppy photo) cannot elevated to the status of significant, the opposite happens; all posts are viewed from a place of "detached reflection".[[41]](#footnote-41) 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 in a way that demands work. 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 only demand an in-platform engagement. In his words: "Nothing is too trivial to be included. Nothing is important that it demands a special place". [[42]](#footnote-42)

Dreyfus cites Kierkegaard as denouncing the press because it makes all information immediately available to everyone. Doing so, it deprives the reader of any notion of what is more local, more relevant, and what is not. Thus, the reader 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".[[43]](#footnote-43)

That being said, we must recognize that information about the trivial, as well as the serious, is a function of social media. One could moderate posts to rid the platform of trivial posts, but this would take away from what social media has to offer now. Therefore, our solution cannot be the strict moderation of all social media platforms.

Instead, we can make use of another type of social media that is rarely a part of the discussion. If we were to bifurcate social media platforms according to how they are organized, Facebook would fall under those that bring users together by common history. On the other end of the spectrum, platforms such as Reddit, Quora and other forums have organizations that make the prior history between users irrelevant, and instead focus on grouping similar types of content. In many of these platforms, user do not even know each other, since the norm is anonymity. Through "threading", which is the categorizing of content depending on its subject matter, these platforms create different local structures with independent communal norms. We can imagine a version of the friend-based social media platforms that adopt threading, so that the claim that "nothing [...] demands its special place" can be superseded.

A similar levelling issue in current friend-based platforms 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. 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.

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.

Furthermore, both platforms have a "suggested friends" bar that appears in the middle of the feeds of both. Thus, the user is continuously encouraged to make connections with new people. Finally, there is no option to keep a post unshared; 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 on the users her post will reach. 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 only tells her closest connections and feels free to use the platform in a more unfiltered way.

Using Sartre's concept, the poster who is not comfortable with the potential audience of her content experiences a peculiar anxiety of being watched by the Other. The Other is a particular individual that observes her judgingly. Recognizing the presence of the Other, the poster can no longer be herself, because she does not experience herself as a subject: "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".[[44]](#footnote-44) 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. Furthermore, The Other is not merely a revelatory power; it changes the subject's being: "All of a sudden, I am conscious of myself as escaping myself, not in that I am the foundation of my own nothingness but in that I have my foundation outside of myself. I am for myself only as I am a pure reference to the Other".[[45]](#footnote-45) Thus, the subject loses her subject-hood, and becomes an object of her own experience[[46]](#footnote-46). No longer a master of her own self, the user experiences what Sartre calls an existential anxiety.

Once again, it is not native to the concept of social media that the poster must feel such an anxiety. The user experiences the look of the Other only when she feels that there is a possibility of another stranger user, for whom the post was not intended, seeing her post. To be fair, the age of the Internet holds that any information that is documented might forever remain on the Internet, but there are ways in which a social media platform could alleviate this anxiety.

Similar to our previous recommendation, we can see a platform that while separating content by subject, 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 gaze and the alienation and mistrust it creates. Different levels of significance and intimacy should be separated by design. As Vaidhyanathan explains: "Different forms of friendship have distinct layers and values embedded in them and operate by different norms".[[47]](#footnote-47) 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 invasion of one's intimacy. By having concretely separated communities with differentiated and explicit norms to which each post is "tagged", the user can choose to view them in mixture or in curation. Having different communities will also make the platform more viable to having pages with ends, so that the experience does not feel infinite, contributing to the "fear of missing our" articulated by Turkle. Furthermore, partially responding to this problem, Instagram recently rolled up a "You are all caught up" feature that allows the users to get a sense of natural conclusion to their time in the platform.[[48]](#footnote-48)

  When there is practically infinite content, curation consumes content. What we experience is no longer a function of what content is available in the planform, but how the platform curates the ordering of the content. Therefore, the question of how platforms curate their content becomes one of the fundamental questions of our experience of social media. With this in mind, the last critique we will consider, articulated by Vaidhyanathan and Zeynep Tufekci, concerns the curation of social media platforms.

The main worry of social media critics on the question of curation is engagement-based news ranking or recommendation algorithms many platforms purportedly use. Vaidhyanathan notes that Facebook -- although the mechanism is generalizable to Instagram and Twitter as well -- orders the posts on a user's home page depending on (along with a few less significant factors) how "engaging" the post is, which is decided on by the indicators of number of "reactions", comments, and shares the post has received so far.[[49]](#footnote-49) The first insight we can construct from this mechanism is that it 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, the post is more likely to get even more attention, which grants it more visibility. Similarly, posts with little engagement early on are less likely to appear towards the top of other users' home pages, which makes them less likely to get engagement later on. In other words, a post "survives" as long as it creates a strong reaction, either partiality or controversy. If it gets neither kind of reaction, it has a short half-life and dies quickly.

From the lurkers' perspective, the clearest side effect of this, which has been discussed mainly in terms of "fake news" and "filter bubbles" but also holds true for all types of posts, is that more emotionally provoking, and consequentially more extremely worded content gets the longest half-life. Although this does not mean that more nuanced posts do not make it in users' home pages, it means they appears less frequently than their more provocative counterparts. Vaidhyanathan says: "most inflammatory material will travel the farthest and the fastest. Sober, measured accounts of the world [will not]".[[50]](#footnote-50) Therefore, users end up getting a picture of their social worlds that is skewed towards more extremism. 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, become the posters' best bet. As the posters' patterns of posting end up validating the lurkers' conception of social media as a trivial platform, they themselves continue the same trend when they are posters, and 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 item[[51]](#footnote-51) that predicts 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 Vaidhyanathan notes, this contributes to the same politically polarizing consequences, such as the formation of filter bubbles: "the intellectual isolation that can occur when 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".[[52]](#footnote-52) Instead of being in a global conversation, we end up creating disjointed circles that always experience the same type of content.

Zeynep Tufekci expresses a similar worry about the "Up Next" column that uses Youtube's recommendation algorithm, which does not create circles among users of the platform but among types of videos. Tufekci claims that looking at a user's watch history, YouTube constructs a personality type for the user shows him videos that "people like him" would watch. This results in an ever-increasing extremism in the types of videos YouTube recommends, not just politically but in all fields. When we watch a video about vegetarianism, we get an "up next" about veganism; "It's like you're never hardcore enough for YouTube".[[53]](#footnote-53) The algorithm pigeonholes the user to a certain type of personality. It is possible that through mere exposure effect, the user ends up fitting his archetype more by watching more of the videos that "his type of a person" would watch, 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" ad-targeting feature, which allows advertisers to use the groupings that the Facebook algorithm makes for curation purposes.[[54]](#footnote-54) The study has found 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.[[55]](#footnote-55) In other words, almost three quarters of political ads were seen by voters who were identified as possibly being convinced by those ads. Less than 20 percent of ads made it to people from both sides of the spectrum close to bilaterally. 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 saw 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.

Another insight we might construct from these facts is that algorithmic curation only allows users to participate in the decision on curation indirectly, from a second-degree. The user gets to pick what he engages in and what he ignores, and in return gets a certain view of his home page the next time he visits the platform. To make a rough analogy, the user's experience is similar to going to a restaurant that brings you a meal depending on your history of enjoyment of previous meals. There is definitely something useful and fascinating in getting a meal that you did not have to particularly order and still enjoyed. However, it rids the user of the possibility of changing preferences whenever he wishes to. If the user of the platform wants to change his experience of the platform, all he can do is dig down and find the posts that he would not normally engage with, engage with them, and continue this intentional pattern-breaking for long enough that the platform notices the change of pattern. But, what if the customer wants to be exposed to different types of food without having to have his taste narrowed down, or worse yet, if he goes vegan overnight? To synthesize the idea with the analogy, an algorithm that makes decisions for us makes it difficult for us to keep an open mind or actively decide to change ourselves. When we are unintentionally pigeonholed by the algorithm, we might get stuck there.

Finally, Tufekci takes caution against the other factors the recommendation algorithms consider that we may not be aware of.[[56]](#footnote-56) She explains that to this day, ML algorithms have been allowed to remain "black box", meaning there has been little effort to structurally make sense of the partial progress in the system. Because of this, algorithms are evaluated not on the basis of whether or not they were able to discover the right characteristics, but on the percent of accurately categorized examples. Therefore, what the algorithm considers is sometimes a mystery. Of course, we will consider the argument in Chapter 3 that human decision making is similarly a mystery; humans are notorious in how biased and inarticulate they are when making decisions. However, Tufekci seems to demand better of machines, given that this is possibly. For computer scientists, the mysterious perception of the algorithms is merely a misperception due to the effectiveness of the algorithms: They can detect patterns that humans cannot, and this is part of what makes them powerful. But the other side of the coin is that even when one wishes to, it is often not possible to figure out what the deciding mechanism of the algorithm is. Tufekci gives the example of targeting plane tickets to Las Vegas: It is uninteresting to us that a company might target single men in the ages of 30-40, because we are aware of this type of strategies, and in some senses, immune to it. However, an ML algorithm might just as well target people with bipolar disorder at the onset of a mania episode. Actually, the latter would be considered a more successful algorithm than the earlier, given its higher percentage of accurate sales. Again, the amorality of targeting a particularly vulnerable group between the two cases differs perhaps in degree, but not in kind. However, implicit in Tufekci's algorithm that since AI-based decision making is not human but human-made, we should demand that it improve on the faults of human decision making.

Whether or not we agree with the claim that these practices are morally problematic, the criticism that engagement-based algorithms encourage radical content and unfairly discriminate against people in targeting ads is one that is widespread in the field of tech sociology. However, among types of criticism we have articulated, this is also the one that is most clearly targeted at a specific feature of the platforms, rather than the medium itself. The 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 set of parameters. 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 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 pure engagement, an alternative social media platform could be protected from these types of criticism.

Furthermore, event current social media platforms have been updating and developing their cost functions to mediate some of the very problems we have described. In January 2019, YouTube has released a statement explaining how they have been combatting videos that have sensational titles but low-quality content, known as "clickbait":

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.[[57]](#footnote-57)

Admittedly, given that Youtube's algorithm is proprietary, we do not have access to the exact nature of this change. However, the declaration at least shows an understanding from Youtube's part that simple parameters such as views do not result in the best user experience, and cost functions generally need to incorporate a set of distinct parameters to combat issues of radical content and model the complex criterion of "viewer satisfaction".

In conclusion, the many blows social media takes from critiques are insightful and necessary for the betterment of the platforms, but they must be viewed as design problems to allow for problem solving to properly function. The questions of notification settings, post structures and curation algorithms all pertain to particular design choices made by the platforms, not mandates of the medium. With this in mind, we use Chapter 3 to anticipate counter arguments to specific claims and to our overarching thesis claim.

**Chapter 3**

In Chapter 2, we claimed that the critique of social media is often misdirected at the nature of the medium, whereas the changes that can amend the states problems are often alternate design choices. One common wisdom of human computer interaction that seemingly challenging software issues can generally be solved by improving the design architecture. Using this methodology, we argued that theorists' critique of social media, alongside being accurate, can be solved through tweaks in design, and we pointed to what these improvements can look like.

We considered the critique in three parts: Notification settings, structure of posted content, and optimization for engagement. In return, we made system-wide recommendations such as specific enhancements to notification settings, the ability to vary, group and tag posted content to allow for more differentiation, and cost functions that optimize more complex notions of user satisfaction than pure engagement. In this chapter, we articulate on what criticism we anticipate, and respond to possible counter arguments against each recommendation we have posited. Some of these arguments are targeted at a specific claim, whereas others target multiple claims in general, or even the overarching argument that the design social media platforms could or should be altered.

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 will always put their own interests ahead of those of the users. In many fields, this criticism undermines productive dialog as to how a system can be improved by rejecting the possibility of change unconditionally. 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.[[58]](#footnote-58)

Vaidhyanathan concludes his analysis on a similarly pessimistic note. Explaining that his critique of Facebook is mostly in vain, he says: "Facebook itself has no incentives to reform".[[59]](#footnote-59) Indeed, one of the theoretical underpinnings of techno-pessimism 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, it is an especially challenging problem to determine who a company should take to be its users. In the example of Facebook, some may claim that Facebook is a platform that is designed for the benefit of the advertisers, and therefore users' dissatisfaction is not unavoidable; it is just not a priority. Of course, in this scenario, Facebook risks the possibility of losing large number of users over time. Similarly, if Facebook were to see its users as its only customers, then we could assume that advertisers would not be satisfied with the platform, and once again, Facebook would be risking losing the business of its advertisers. Therefore, without knowing the strategic development decisions that go into any proprietary platform, we can assume that it would be beneficial for them to consider both the users and the advertisers as their customers, since without either of the two, the platform is not profitable. Hence, the question of business model is not a stand-alone; it only makes sense coupled with the question of customer base.

There is empirical evidence to support the theoretical claim that platforms should care about user satisfaction. 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.[[60]](#footnote-60)

Of course, Vaidhyanathan or any techno-pessimist can point out that words are free. However, failing to respond to user demand is not. An example of radical change in business models that is relevant today is the rise of what is called "fast-casual" food, and the fast adaptation of the largest fast-food chains to join the trend. In the last couple of years, McDonald's removed problematic ingredients from its menu, such as certain artificial preservatives; Taco Bell made a claim to become one of the healthiest fast-food chains;[[61]](#footnote-61) and the whole industry reduced the average calories of a menu item by 12% in one year.[[62]](#footnote-62) The drastic changes, contrasted with earlier models that fast-food chains have incorporated, are hard to explain without considering the rise of a health-conscious customer base.

Numerous studies that are replicated frequently, including those from CMU and Stanford, showcase the significantly negative social and psychological effects of social media use as it is experienced today.[[63]](#footnote-63) 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 perception will catch up eventually. The fast food industry, after decades of making food more chemically-developed, more tasty but unhealthy, finally went through this change as the tremendous increase in diabetes rates finally caught the public eye.

Today, existing 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.[[64]](#footnote-64) Therefore, an alternative design is not only realistic, but perhaps also reasonable. How a platform (or even a new medium) started 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 unreasonable to demand that car manufacturers design all cars with this invariant.

Another counter argument we anticipate is one that pertains to a number of the claims we have posited. In Chapter 2, we claimed that social media platforms inadvertently favor and encourage radical content, while discouraging nuanced or calculated content. We suggested many root causes for this problem, such as the invariant structure of posts that causes users to give brief and shallow attention to any given post, the infinite scroll-feed that gives the user no natural stopping point, cost functions that optimize for engagement and encourage users to provoke their audiences, etc. Since this claim grounded many of the problems and suggestions for which we argued, we anticipate a counter argument that trivializes our claim that the dissipation of radical content is a legitimate problem.

In short, we anticipate the argument that radical posts make up a small portion of social media posts. It might suggest that perhaps radical posts are "louder", but they are problematic in effect only because they are brought under the spotlight by media outlets or other critics. In this way, perhaps radical posts are only as big as a problem as we make of them, and if we left them alone, they would be contained in a small area of social media platforms from which one can easily distance herself.

Although this account is definitely optimistic and convenient, it is nonetheless not based on truth. First of all, 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. [[65]](#footnote-65)

In other words, no matter how the content regulation rules are set up, users tend towards posts that are on the edge of the regulations. From Zuckerberg's perspective, this is not a property of the system, but a property of human beings. Indeed, there is a vast body of research that supports Zuckerberg's claim that humans are attracted to the sensationalist, edgy and radical content. Even directly concerning news content, studies show that the prioritization of the sensational is nothing that began with Facebook. 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.[[66]](#footnote-66) Studies also show that attraction to gossip, scandals, and other types of content that create emotional response has been invariant throughout human history; some even suggest that gossiping about the scandalous is an evolutionary trait that humans have developed to enhance social bonding through dialog.[[67]](#footnote-67) Therefore, in the case of human communication in general, and of social media content in specific, we can safely assume that regardless of its size, whatever content sparks emotional response will get more attention. Unfortunately, we cannot avoid the problem of edgy content by simply trivializing its size.

Actually, the radical content must be problematized in another prospect as well. Due to the network-like structure of social media, the issue of the prevalence of radical content is exacerbated even further: First, radical posts get unproportionate engagement due to users' inherent interest in them, and second, their distribution increases non-linearly due to what is called the "network effect". 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.[[68]](#footnote-68) In other words, things grow in networks at a higher rate than outside of networks. The network's effect is that it enhances the growth of its content.

In the case of social media, Metcalfe's law says that whatever gains attention will gain even more attention. This is a crude conceptual ground for why content gaining early recognition more frequently "goes viral" than content that doesn't catch an early wind. If a post successfully beings an initial dissipation, it is more likely to spread similar to a virus. Indeed, the spread of virus itself is object to a network effect as well, since human cells, and computers, are connected through network systems. Metcalfe defines his law in terms of "value" and claims that a network's value is non-linear in its size. However, in the case of radical posts in social media, we claim that the non-linear dissipation is not added value but added stress to the system. Therefore, the social media platform must be sensitive to the network effect and given our claim that the dissipation of radical content hurts the system, alter its algorithm to ameliorate the network effect radical content enjoys.

Another counter argument that is connected to the claim on optimization algorithms suggests that it is unproductive to expect social media platforms to simply produce "fairer" algorithms and fix problems of content moderation and unfair ad targeting completely. It will become clear later that this argument is one that we agree with whole-heartedly, and our recommendations are ones devised precisely out of this concern. However, we will begin by detailing the argument.

In their paper *Inherent Trade-Offs in the Fair Determination of Risk Scores,* 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.[[69]](#footnote-69) 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. Of course, one could argue that the paper might have chosen one or more of the notions of fairness to be unreasonable ones, which makes the finding ineffective, even if accurate. However, the authors suggest and cite cases in which all three of the properties have been used as significant in judging an AI system's fairness. Furthermore, when we consider the three properties, we can see that they go hand in hand with our intuitive understanding of fairness and don't seem to contradict one another from the outset. The three properties, in short, say: First, scores (outputs) of the algorithm should have the same meaning regardless 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.

Although the paper is written mainly for risk assessment systems, we can extrapolate what these "classes" would correspond to in news ranking or targeted ads. One such example from Chapter 2 is Tufekci's argument that an individual with onset of a mania period might be targeted because he is more likely to make impulsive purchases. Since she finds this unfair, we can say in this situation that this individual would be in the "negative class". Similarly, we can make such groupings for individuals with certain political affiliations, racial groups, etc. Therefore, their conclusion that these properties of a fair result cannot be simultaneously guaranteed is a powerful conclusion. Indeed, in social discourse, we generally take an agreement on definitions to be a necessary, if not sufficient, requirement in having a productive discussion. What the authors demonstrate here is that we do not yet have a common ground understanding of what fairness is, and in fact, we have definitions that yield incompatible results.

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 in-practice status quo, which in the case of decision making, is human beings. Therefore, we must question how reliable humans are as judges when making claims about AI judgment. 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.[[70]](#footnote-70) 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 probably much higher than in a social media recommendation algorithm. 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 an outcome of a number of independent trials that has been produced few times is more likely to be produced in the next trial.[[71]](#footnote-71) Finally, numerous studies tell us that in many fields, AI decision making is more unbiased than human decision making.[[72]](#footnote-72) Therefore, from a design point of view, the argument for ideal algorithms may be too high of a bar, when the status quo is not that ideal to begin with.

In Chapter 2, we suggested that there are problems of disempowerment due to the lack of understanding and control users have over ML algorithms used in social media platforms, mostly in news ranking and recommendation systems. However, it seems that an approach to make algorithms fairer is unproductive, since fairness is admittedly not a well-defined term, and comparing fairness of algorithms with that of our ideal does not hold, since human judgement, which is the incumbent judge in making decisions about how the platform should look to users, are infamously biased in all areas of life, making the comparison to ideals an unrealistic one. That being said, we cannot use this critique to deem the problem impossible and settle with the status quo. The recommendation that users have more control over and understanding of the algorithms that govern them does not necessitate that software engineers be mandated to develop unflawed code. Even if making uncontested progress in ML-judgement is difficult, we can still create platforms that are more explainable so that users can be aware of the considerations that go into the decision-making process, and even have a way of influencing the algorithms by changing their history.

We have in our capacity today the ability give information to the user about what parameters were considered for a specific decision to be made. In other words, we can challenge the supposed "Blackbox" nature of the algorithms. In their paper titled *Accountability of AI Under the Law: The Role of Explanation,* Finale Doshi-Velez et al tell usthat we could reasonably expect as much transparency in decision making of algorithms as we do from humans.[[73]](#footnote-73) In this work, the authors ask how we could continue to take advantage of the power of AI systems while still holding them accountable, and they respond: "explanation".

Taking the precedents of legal perspectives as what explanations we accept as valid in human judgment, they suggest that the same form of explanation can be expected of AI algorithms as well. The two methods they employ are "local explanation" and "constructing counterfactual scenarios". To understand these tools, first we must recognize that explanation and transparency are not equivalent. For example, a social media platform could give out their source code, but this would not necessarily 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. With this in mind, 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?".

A local explanation is an "explanation for a specific decision, rather than an explanation of the system’s behavior overall". Even though algorithms look for an abundance of parameters and give them weights through non-linear (and therefore difficult to calculate) functions, the study shows that the algorithms could be questioned about particular decisions in these human-friendly ways. Similarly, constructing counterfactual scenarios is the algorithm's ability to respond to the question: "Would changing a certain factor have changed the decision?" By simulating the algorithm with an altered particular parameter, it could tell us whether this 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".

Without much difficulty, we can imagine how designing a platform with such explanations regarding its news ranking or suggestions algorithm can help with the problems of disempowerment we described in Chapter 2. First of all, we explained that users can get inadvertently pigeonholed into a category which they are not happy with. For example, we can 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". With an algorithm that allows for explanations, Tufekci could ask whether a particular video would be in the Up Next if she had not seen another particular 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.[[74]](#footnote-74) In a platform with explanations, Tufekci could identify the videos that have pigeonholed her into a certain category and remove them from her watch history. Similarly, Mark Zuckerberg recently claimed that they are 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.[[75]](#footnote-75)

At the same time, the explanation allows the users to have they type of first-order-input that we suggested is lacking in current systems. By tweaking their user history to determine which behavior they would like to affect the recommendation algorithms and news ranks, users get to make explicit decisions about their own experience of the platform, thus being empowered in a way that they currently are not. 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, the platform could still grant the users the ability to change the system's behavior in predictable and useful ways.

However, one drawback of exposing any information in a proprietary system is that it introduces a vulnerability into the system, namely that users of the system are better able to manipulate the system for a result desired to them. In the case of social media users, we might be tempted to claim that it should be a feature that users can get the desired results, since that is what we have been advocating for. However, for any given system, there is an underlying threat model, and in any threat model, there are actors who wish to abuse the system. Without dwelling too much on the specifics, we can speculate that the threat actors for a social media platform can 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. Hopefully, local explanations of the sort that Doshi-Velez describe are user and instance-specific enough that they will not enhance threat actors' understanding of the underlying algorithm in any meaningful way. However, it must be noted that there is an inherent tradeoff of exposing information about a system, and this tradeoff requires careful deliberation from the side of the developer.

Finally, the two main counter-arguments that pertain to a specific feature we have recommended are those about the enhancement of notification settings. The first problem we articulated in Chapter 2 was Turkle's "life-mix", the co-presence and conflation of the individual's physical presence with his online presence. We found that the main bridge between those two selves is the red alert notification, the invitation that calls the user to move from the physical world to the online world. Rather than study the fact of a social media platform having notifications at all, which has many advantages, such as the ability to alert the user to relevant activity, we questioned how the notifications in current platforms behave. For this, we turned our attention to notification settings.

We suggested two improvements on notification settings in an alternate design. First, we recommended that the default notification alerts, both when the user signs up to the platform for the first time, and as new notification alerts emerge, should not be completely opt-out. Further, we recommended that the user be directed to the notification settings page when she first signs up, so that she can alter them to her liking at that moment and be aware of the altercations she is allowed to make for future use. Secondly, we suggested that users have a way of dealing with notification settings as a whole, such as the ability to turn on a "do not disturb" mode, which pauses all notifications until a desired date. In response to these recommendations, we expect two counter arguments.

The first counter argument is that human beings are logical creatures, so they should not have any problems identifying what notification settings they would like to have, finding the corresponding settings, and changing them. In other words, 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.

Nobel Award recipient behavioral economist Richard Thaler names the predictable ways in which human behavior deviates from the dictates of logic as "nudge theory". The mantra of this theory has been summarized as: "If you want people to do something, make it easy".[[76]](#footnote-76) In other words, even if it there is a set of decisions that are more advantageous for a person, she may fail to make the "rational" choice and settle with the default if there are barriers to changing behavior, no matter how small. 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. Opt out rates average about 10%".[[77]](#footnote-77) The study eliminates other possible explanations to be able to make the claim that the increase is tied to the default option. If individuals are unwilling to a significant extent 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 anything that affects how difficult it is to make a change. Thaler's paper outlines other results that reaffirm the potential impact of our 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.[[78]](#footnote-78) Then, we can safely 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 second counter-argument we anticipate is that providing more options cannot be a blanket-solution to all software problems, since this itself introduces new problems of user-friendliness. If there was no tradeoff, social media platforms (and any other technology, for that matter) could introduce more granular preferences ad infinitum. For notification settings, as well as for other areas of social media design, we have made recommendations that include adding more options. Therefore, this criticism deserves recognition and response.

Hick's law suggests that as the number of options in a given environment increase, the time it takes for a human being to decide on what to do next increases.[[79]](#footnote-79) Although there are exceptions to this, it is true that 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 combined with another fact: Human attention is a scarce and fragile resource that is easily distracted, especially when challenged.[[80]](#footnote-80) 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 in which the user could easily fall on the default settings in order not to make decisions that challenge their 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, thus failing to have the impact the choices were intended for. 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".[[81]](#footnote-81) If the decision is necessary 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 we must recognize. Design is ultimately a game of finding the right balance between tradeoffs; compromises must be made to ensure a happy medium between complexity and simplicity; between flexibility and usability. Therefore, we turn to the suggestions of design literature on how to mitigate the complexity added by 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. Breaking down the decision process such that going into each screen, the user knows what choice they make allows us the make use of the main exception to Hick's law: If the user knows what they want before encountering the screen, having more choices does not have the same effect. In this case, the user simply peruses until she finds the option she is searching for, rather than deliberate on each individual option. Hence, our recommendation in Chapter 2 includes a set of meta-options: "Permissive", "Conservative", and "Power User". By grouping all notification settings in three broad categories, we hope to avoid overwhelming the user without losing functionality.

Furthermore, by grouping the items in these specific categories, we make 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 UX, it argues that 80 percent of users use merely 20 percent of the functionality.[[82]](#footnote-82) In the case of notification settings, we assume that the needs of most users are covered by either permissive settings, which basically correspond to the default settings social media platforms provide today, or conservative settings, which require opt-in consent from users who wish to be mostly unbothered by notifications. The power user case allows us to target the remaining 20 percent of the users who wish to build custom settings for each type of notification.

Thus, we conclude our discussion of anticipated counter-arguments, and turn to a summary of the farther-reaching implications of the design choices we have recommended.

**Chapter 4: Conclusion**

Throughout history, the main medium of mass communication of the day has been held to scrutiny, be it the newspaper, the radio, the television, or the "global village". It is famously known that Socrates was against writing, because the medium would destroy authentic dialogue, threaten the intimacy and the soul-to-soul nature of communication, and provide "not truth, but only the semblance of truth".[[83]](#footnote-83) Yet, tools of communication continued to develop, and although some forms of critique have remained constant, each medium has mostly been questioned and challenged according to its particular features, and its unique situatedness in a specific time and space.

Today, social media is uncontestably one of the most popular forms of mass communication. As expected, thinkers challenge every aspect of the medium, and some find no significant redeeming qualities in it. There are different forms this argument takes. In the field of political science, Vaidhyanathan's *Antisocial Media* claims that the problem with Facebook "is Facebook", and finds no way out of the irreparable damage it claims Facebook has caused, except for the closing down of the platform. In philosophy, Dreyfus suggests that the nihilism Internet has successfully harbored can only be avoided if one turns off the monitor and finds a particular hobby or cause that will guide their human endeavors. In sociology, Tufekci likens the ad-based platform to an AI-powered dystopia. These thinkers are not radicals of a generally-moderate group; they are the authors of some of the most well-articulated and sound critique social media faces today. Thus, even if we do not accept these critiques as fact, we must admit that they are insightful problematizations that need to be reformed.

However, the techno-pessimism that we tend to slip into especially when we are highlighting the most problematic aspects of social media is not only an unproductive approach to pave the way forward, but also misdirected. The medium of social media does not require that content be categorized by posts of a certain format, distributed in a specific algorithm, or accessed through a particular interface. These are decisions that creators of today's platforms have made; decisions that they continue to make on a daily basis. At its genesis, Facebook did not have a timeline, and the home page only displayed activity the user herself was involved in. Instagram only established "stories" in 2016, after a completely separate platform was built around the idea. Twitter famously started with a 140-character limit, which was doubled as it became obvious that users built ad hoc solutions to get over the limit, such as numbering their posts so that they are read as a single piece. Some of these changes may seem trivial to the reader, but the point is that social media platforms are not constants we have to either accept or reject from the outset; they are subject to change and influence through incentives.

Therefore, the ambition of redirecting the criticism considered in this thesis 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, or acceptance, of any system disallows progress and hinders advancement, replacing these with the singular aim of the disposal, or an indiscriminate spread, of the system. Uncritical acceptance of techno-pessimism, as well as tecno-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 our tasks to evidence and advocate for the insightfulness of their comments. However, if we allow the first-generation critique against social media to remain as pointed as it has been, we would be encouraging the aforementioned radical and unproductive positions. As the next generation of work that follows this grand scale critique, we must further it in the direction of more specificity, more nuance and more individuation. Thus, our primary goal has been to save social media from critique that is better understood if 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. Furthermore, even after the user enters the platform in such non-ideal settings, the sizable world inside the platform provides the users with its own discontents. The structure in which content is presented in these platforms falls contrary to how we as humans make the world intelligible to ourselves. In return, we encounter the 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 thought, as well as the impairing of exploratory attitudes, replacing them instead by narrow and confined echo chambers.

The sheltering of the medium from the criticism is not meant to encourage a policy of inaction towards social media. To the contrary, it is meant to find productive paths towards how to improve on the platforms without giving up on the ideal of global communication. Therefore, our recommendations are concrete, even if this requires giving up on grandiose statements and instead focusing our attention on minute details of the systems. 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 the platforms and allowing independent communal norms, as well as a variety in the design of the publishing of content. Finally, the radicalization and pigeonholing of the content the user encounters and feels the need to partake in can be retaliated by incorporating particular parameters into loss functions to optimize for.

Each claim made in this work is subject to criticism. Neither the mapping of the critique to design features, nor the proposed alternative architectures are 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, we propose this piece of work 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 nature 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 whole way 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 mass communication, and pave the way towards the redemption of a now-trivialized and contaminated ideal of global interconnectedness.

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. [↑](#footnote-ref-1)
2. Dreyfus, Hubert L. *On the internet*. Routledge, 2013. [↑](#footnote-ref-2)
3. Heidegger, Martin. “The Thing.” *Poetry, Language, Thought.*, Harper & Row, 2001. [↑](#footnote-ref-3)
4. Dreyfus, *OI*. [↑](#footnote-ref-4)
5. Postman, Neil. *Technopoly: The Surrender of Culture to Technology*. Vintage Books, 1993. [↑](#footnote-ref-5)
6. Dreyfus, *OI*. [↑](#footnote-ref-6)
7. Gershon, Ilana. "Un-friend my heart: Facebook, promiscuity, and heartbreak in a neoliberal age." *Anthropological Quarterly*84, no. 4 (2011): 865-894. [↑](#footnote-ref-7)
8. Tufekci, Zeynep. "We’re building a dystopia just to make people click on ads." *TED: Ideas Worth Spreading, Sept* (2017). [↑](#footnote-ref-8)
9. Vaidhyanathan, Siva. *Antisocial media: How Facebook disconnects us and undermines democracy*. Oxford University Press, 2018. [↑](#footnote-ref-9)
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/. [↑](#footnote-ref-10)
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. [↑](#footnote-ref-11)
12. Zittrain, Jonathan. *The Future of the Internet--and How to Stop It*. Yale University Press, 2009. [↑](#footnote-ref-12)
13. McLuhan, Marshall. *Understanding Media; the Extension of Man*. McGraw-Hill Book Company, 1965. [↑](#footnote-ref-13)
14. "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. [↑](#footnote-ref-14)
15. McLuhan, Marshall, and Quentin Fiore. *The Medium Is the Massage: an Inventory of Effects*. Penguin, 2008. [↑](#footnote-ref-15)
16. Turkle, Sherry. *Alone Together: Why We Expect More from Technology and Less from Each Other*. Basic Books, 2017. 329. [↑](#footnote-ref-16)
17. Ibid, 155. [↑](#footnote-ref-17)
18. Ibid, 160. [↑](#footnote-ref-18)
19. Ibid, 161. [↑](#footnote-ref-19)
20. Ibid, 163. [↑](#footnote-ref-20)
21. Ibid, 164. [↑](#footnote-ref-21)
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.

    [↑](#footnote-ref-22)
23. "Notification Basics & Settings | Facebook Help Center." Facebook. Accessed March 26, 2019. https://www.facebook.com/help/327994277286267. [↑](#footnote-ref-23)
24. Ibid. [↑](#footnote-ref-24)
25. Facebook. Accessed March 26, 2019. https://www.facebook.com/settings?tab=notifications§ion=on\_facebook&view. [↑](#footnote-ref-25)
26. Ibid. [↑](#footnote-ref-26)
27. "Login Button - Facebook Login - Documentation." Facebook for Developers. Accessed March 26, 2019. https://developers.facebook.com/docs/facebook-login/web/login-button/. [↑](#footnote-ref-27)
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. [↑](#footnote-ref-28)
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. [↑](#footnote-ref-29)
30. Ibid. [↑](#footnote-ref-30)
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. [↑](#footnote-ref-31)
32. Ibid. [↑](#footnote-ref-32)
33. Vaidhyanathan, *AM*, 1. [↑](#footnote-ref-33)
34. Ibid, 4. [↑](#footnote-ref-34)
35. Ibid, 5. [↑](#footnote-ref-35)
36. V Shred's Facebook Page, Accessed March 26, 2019. https://www.facebook.com/watch/?v=2267436653579031

    [↑](#footnote-ref-36)
37. 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. [↑](#footnote-ref-37)
38. 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. [↑](#footnote-ref-38)
39. Vaidhyanathan, *AM,* 17*.* [↑](#footnote-ref-39)
40. Dreyfus, *OI,* 75. [↑](#footnote-ref-40)
41. Ibid, 72. [↑](#footnote-ref-41)
42. Ibid, 79. [↑](#footnote-ref-42)
43. Ibid, 76. [↑](#footnote-ref-43)
44. Sartre, Jean-Paul. "The Look". *Being and nothingness*. Open Road Media, 2012. [↑](#footnote-ref-44)
45. Ibid. [↑](#footnote-ref-45)
46. Ibid. [↑](#footnote-ref-46)
47. Vaidhyanathan, *AM,* 47. [↑](#footnote-ref-47)
48. "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/. [↑](#footnote-ref-48)
49. Vaidhyanathan, *AM,* 6. [↑](#footnote-ref-49)
50. Ibid. [↑](#footnote-ref-50)
51. Ibid. [↑](#footnote-ref-51)
52. "What Is a Filter Bubble? - Definition from Techopedia." Techopedia.com. Accessed March 26, 2019. https://www.techopedia.com/definition/28556/filter-bubble. [↑](#footnote-ref-52)
53. Tufekci, *Dystopia.*  [↑](#footnote-ref-53)
54. Merrill, Jeremy B. *Targeted Facebook Ads*. [↑](#footnote-ref-54)
55. Ibid. [↑](#footnote-ref-55)
56. Tufekci, Zeynep. "Machine intelligence makes human morals more important." *TED Talk* (2016). [↑](#footnote-ref-56)
57. "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. [↑](#footnote-ref-57)
58. 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/. [↑](#footnote-ref-58)
59. Vaidhyanathan, *AM,* 17. [↑](#footnote-ref-59)
60. 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/. [↑](#footnote-ref-60)
61. 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. [↑](#footnote-ref-61)
62. 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. [↑](#footnote-ref-62)
63. Dreyfus, *OI.*  [↑](#footnote-ref-63)
64. 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/facebooks-scandals-in-2018-effect-on-stock.html. [↑](#footnote-ref-64)
65. Constine, Josh, *Zuckerberg Highlights.*  [↑](#footnote-ref-65)
66. Davis, Hank, and S. Lyndsay McLeod. "Why humans value sensational news: An evolutionary perspective." *Evolution and Human Behavior* 24, no. 3 (2003): 208-216. [↑](#footnote-ref-66)
67. McAndrew, Frank T. "Can gossip be good?." *Scientific American Mind* 19, no. 5 (2008): 26-33. [↑](#footnote-ref-67)
68. 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. [↑](#footnote-ref-68)
69. Kleinberg, Jon, Sendhil Mullainathan, and Manish Raghavan. "Inherent trade-offs in the fair determination of risk scores." *arXiv preprint arXiv:1609.05807* (2016). [↑](#footnote-ref-69)
70. 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. [↑](#footnote-ref-70)
71. 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. [↑](#footnote-ref-71)
72. "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. [↑](#footnote-ref-72)
73. 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).

    [↑](#footnote-ref-73)
74. 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/. [↑](#footnote-ref-74)
75. Constine, Josh, *Zuckerberg Highlights.* [↑](#footnote-ref-75)
76. "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/. [↑](#footnote-ref-76)
77. Benartzi, S., & Thaler, R. (2013). Economics. Behavioral economics and the retirement savings crisis. *Science (New York, N.Y.),* *339*(6124), 1152-3. [↑](#footnote-ref-77)
78. Ibid. [↑](#footnote-ref-78)
79. The Interaction Design Foundation, *The Pareto Principle.* [↑](#footnote-ref-79)
80. 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. [↑](#footnote-ref-80)
81. The Interaction Design Foundation, *The Pareto Principle.* [↑](#footnote-ref-81)
82. Ibid. [↑](#footnote-ref-82)
83. Plato, C. J. "Phaedrus." (1988). [↑](#footnote-ref-83)