Date: February 2021

Performing the Platform: A Taxonomy of Content Manipulation

By Holly Robbins for Bits of Freedom.

TABLE OF CONTENTS

INTRODUCTION

Your social media feed is not a mirror, reflecting the world exactly as it is back to you. Rather, it is the result of numerous mechanisms that have been designed to serve the objectives of the companies operating the platforms we use to share and receive information.

As the world grapples with these new digital agoras for publics to congregate and create discourse, there will be an inevitable period of trial and error to determine what shape and role these platforms should have in our society. The prevalence of these platforms, and their capacities to shape our realities, demands that we critically engage with how these platforms are designed. Specifically, this report scrutinizes how the design of these platforms seeks to mold user behaviors and engagement with the platform. This molding is a form of manipulation.

Through design analysis, reporting and academic research, this report identifies and labels six unique forms of manipulation carried out by Facebook, Instagram and YouTube. It describes the various design decisions that constitute these forms of manipulation, and discusses their implications. With this taxonomy, we hope to contribute to increased "platform literacy". If we are better able to recognize, name, and understand the types of manipulation and their implications, we will be better equipped to shape the future and role that these technologies play in our society.

i. Definition

In the context of this report, manipulation is examined in terms of the relations that exist between the manipulator and the manipulee. In this case: "manipulation" is when a manipulator steers or controls the manipulee (typically covertly),[1] with the aim of making the manipulee a part of a self-serving scheme of the manipulator, with little to no concern for the wellbeing of the manipulee.[2]

"Content manipulation" refers to how content, including users, on social media platforms is steered or controlled by the platform with the (likely) intention of steering a person to engage in a scheme that benefits the platform. Content manipulation is carried out through various means and is motivated by different economic, political, and social forces and informs almost every design decision made by Facebook, Instagram, and

1 This is what distinguishes "manipulation" from "persuasion." See: "Technology, autonomy, and manipulation" by Susser, Roessler and Nissenbaum.
2 *Ibid*.

operate, to the opportunities that are created for users to engage with the platform.

ii. Scope

Today's social media platforms serve as a contemporary agora for public discourse, as well as being essential infrastructure for interpersonal communication. This report examines content manipulation on the three most popular social media platforms in the Netherlands: Facebook, Instagram and YouTube. This report does not offer an exhaustive account of particular manipulative tactics, software and design, but instead highlights trends and patterns that demand critical engagement.

iii. Taxonomy of content manipulation

This report describes six forms of content manipulation. Each type represents a different tactic or approach used by the platform to direct the user to take part in some type of self-serving scheme of the platform itself. Table 1 on page 3 identifies and defines these different forms of manipulation, as well as lists some of the key mechanisms and concepts that support these specific forms of manipulation.

The consequences of these forms of content manipulation do not only reside online, but translate to very real offline consequences as well. For example, the spread of misinformation that fuels the outcomes of significant political actions (chapter 4), or the rising levels of addiction and dependency (chapter 5), to name a couple. Part of what makes these forms of content manipulation so successful is the lack of transparency and absence of accountability. These businesses are born from preexisting archetypes (newspapers, business models, etc), yet have evolved into a new category of organization that existing legislative and regulatory precedents haven't yet been fully capable of accommodating.

Type of manipulation	Definition	Key Mechanisms
Micro-Manipulation	Design interventions at the level of the user interface that direct user behavior according to the aims of the platform.	Dark Patterns, shortcuts, interface design
Content Moderation	The system by which a platform identifies, classifies, and permits or removes user content, and the mechanisms it offers through which users can engage with this process..	Content moderators, automated filters, community guidelines and standards
Algorithmic Content Curation	Design interventions at the level of the software that determin what (types of) content should be made more or less visible on the platform, and to whom.	Algorithmic feed, engagement metrics
Micro-Targeting	The system by which user behavior is collected, analyzed and used to generate a profile of the user, which in turn is mechanized to manipulate the user.	Data collection, profiling, content discrimination.
Psycho-social Manipulation	Interventions aimed at creating an addictive environment for users playing to their vulnerabilities and psychological needs.	Validation, variable-ratio reward schedule
Self-Manipulation	When the user manipulates their own behavior and content on these platforms to conform to certain expectations (projected or otherwise) of the platform.	Shadow banning, influencing, “hacking” the algorithm

1. MICRO-MANIPULATION

When designers create websites, they must try to predict how to best accommodate the user's wants and needs. For example, how can the interface help the user find a particular navigation menu, or a short cut to another service on the site? Broadly speaking, this is the objective behind interaction design: to use design as a means to support or enable certain types of user behavior or interactions.

This report argues that these design tactics represent a form of manipulation because they solicit a particular type of response or behavior from users (the manipulated) to serve the purposes of the platform (the manipulator). These are small acts of manipulation (as opposed to larger systematic design choices embedded in the platform, which will be discussed later) that target individuals, thus we categorize these as "micro-manipulation."

This chapter will first discuss particularly nefarious design tactics that appear on the screen known as "dark patterns." Dark patterns are design tactics that are specifically meant to "trick" a user to do something that they may not have otherwise intended to do. Later, this chapter discusses design tactics that are not necessarily intended to "trick" a user, but are more general design choices that shape the type of service and forms of communication that the platform offers. These more general design choices illuminate the agenda and priorities for the types of user behavior that the site encourages.

1.1 Dark patterns

Design interventions to direct user behavior can be made with the best intention of the user in mind (helping the user to find the log-off button). However, these designs may also try to manipulate, or "trick," the user to do things that they didn't intend to do, such as signing up or buying something, or making it incredibly complicated to unsubscribe from a service. In these cases, design interventions are not intended to serve the needs of the user, but those of the platform, website or service itself. This is what represents the fine line between nudging[3] and manipulation. These misleading designs that act to direct user behavior in a way that is not be in the best interest of the user are referred to as "dark patterns."

3 "Nudging" is a concept from the fields of psychology and behavioral economics that is concerned with how users can be lead towards making certain choices by appealing to their psychological raises. For example, setting your gym cloths out on a chair at night to help encourage yourself to put them on and work out first thing in the morning.

Dark patterns are a well-documented design tactic. Typologies and categories of dark patterns have been made by academics and design practitioners alike.[4] This report will not provide an exhaustive account of the dark patterns used on YouTube, Instagram and Facebook. For one, these platforms are constantly being updated and redesigns happen so regularly that attempting to capture instances of these design tactics would be futile.[5] Instead, this section will capture examples of dark patterns and offer some context as to their implications, as well as help people develop a critical eye to be able to spot how certain interface design may be attempting to "trick" them into a particular action.

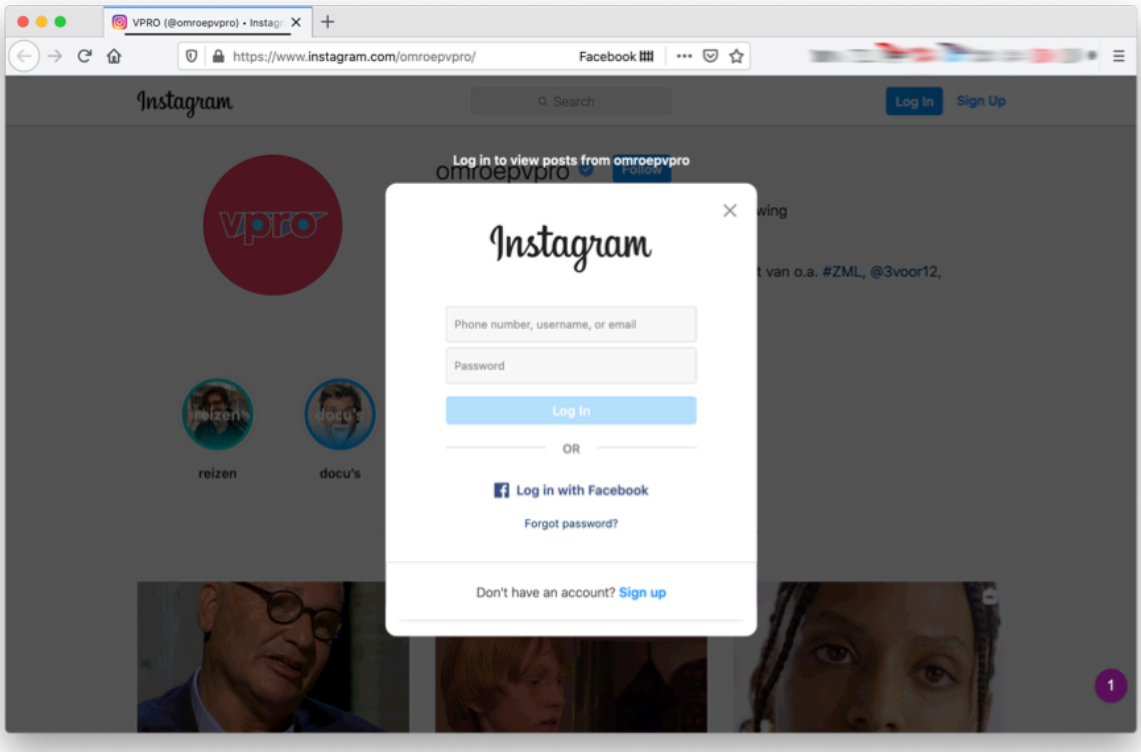
1.1.1 Forced Action

"Forced Action"[6] is a design tactic where users are required to perform a specific action to use the platform. For example, a user cannot view (certain) content on Instagram until they themselves have an account and are signed into that account. These can also come in the form of cornering mechanisms, such as an agreement that pops up once a new program has been launched or a new update has been downloaded. The menu forces the user to agree to the updated agreement before being permitted to continue to use the platform. These notifications become a barrier between the user and the task they came to the platform to perform.

4 The typologies of dark patterns featured in this report are an amalgamation of several different resources. But two primary resources informed this section: For a more analytic framing of categories of dark patterns, see: "The Dark (Patterns) Side of UX Design" by Gray, Kou, Battles, Hoggat and Toombs, and see: Darkpatterns.org for more specific subgroupings of typologies. For a discussion on how design is being used deceptively to regulate privacy settings (specially on Facebook and Google) see: "Deceived by Design". For a more specific perspective on how they are implemented on e-commerce sites, see: https://webtransparency.cs.princeton.edu/dark-patterns/.

5 Darkpatterns.org is a good resource for novel trends in dark patterns (on Facebook, Instagram, YouTube, and beyond).

6 "The Dark (Patterns) Side of UX Design".



1.1.2 Social pyramid

Platforms or services incentivize users to recruit other users to that service in a manner that can resemble a pyramid scheme. For example, Facebook and Instagram request users to synchronize their address book with their accounts in an effort to identify more accounts to follow. Likewise, those contacts are notified of the new user on the platform and are encouraged to follow that user. This has the impact of creating more opportunities for users to grow their feeds and the time they spend online.[7] Another example can be found in the popular Facebook game FarmVille, which offered users features or specialized goals on the condition that they invite their friends to join the game.

This tactic isn't only deployed by the platforms and directed at users. Perhaps more nefariously, the design of the platform can incentivize people to use this social pyramid tactic on each other.[8] For example, on Instagram, the metric of success comes in the form of engagement with content. Therefore, businesses on Instagram are motivated to develop incentives for people to engage with their content so that that very content might be made more visible to followers of their followers and

7 Chapter 3 discusses how this particular tactic is economically motivated.

8 This is described in more depth in chapter 6.

perhaps be featured more prominently on the feeds of new, potential followers. In this case, followers may be requested to tag their friends in the original post, or promote the original content in their own feeds, in exchange for rewards or benefits.

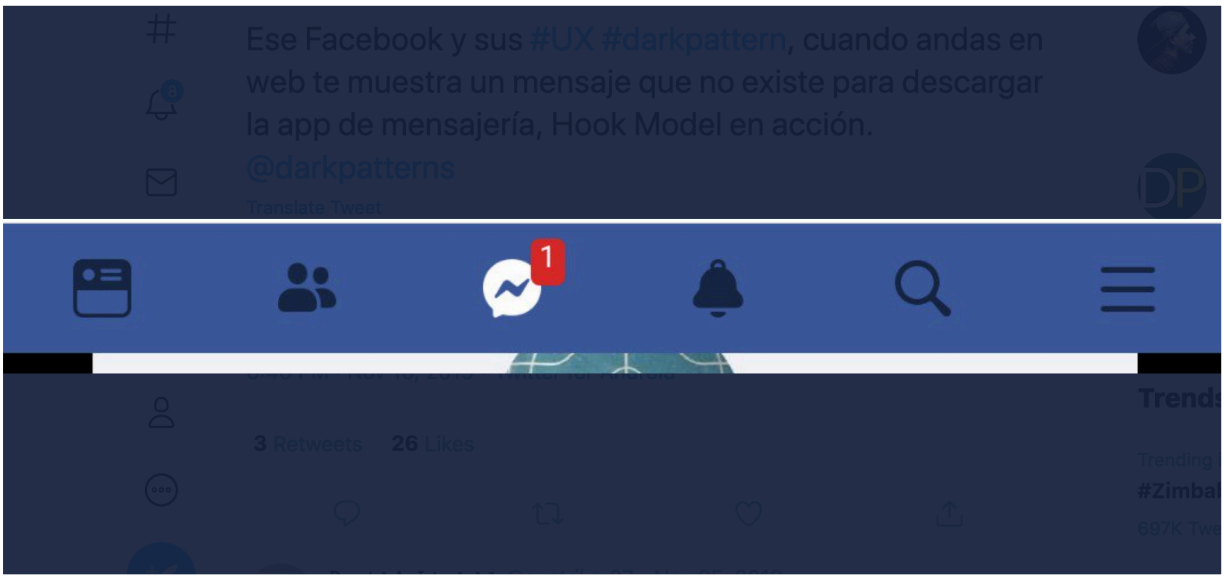
Perhaps the most well known example of a social pyramid scheme is held into place by a lack of interoperability between services. This leads to the creation of walled gardens that force people to use the service their friends, family or co-workers use: the network effect.

1.1.3 Sneaking

Sneaking is a category of design tactics that attempt to hide, disguise, or delay the divulging of information that is relevant to the user. This occurs among others in the surreptitious ways that platforms bury information about how they may make use of user data in lengthy and dense terms and service agreements.

1.1.4 Bait and switch

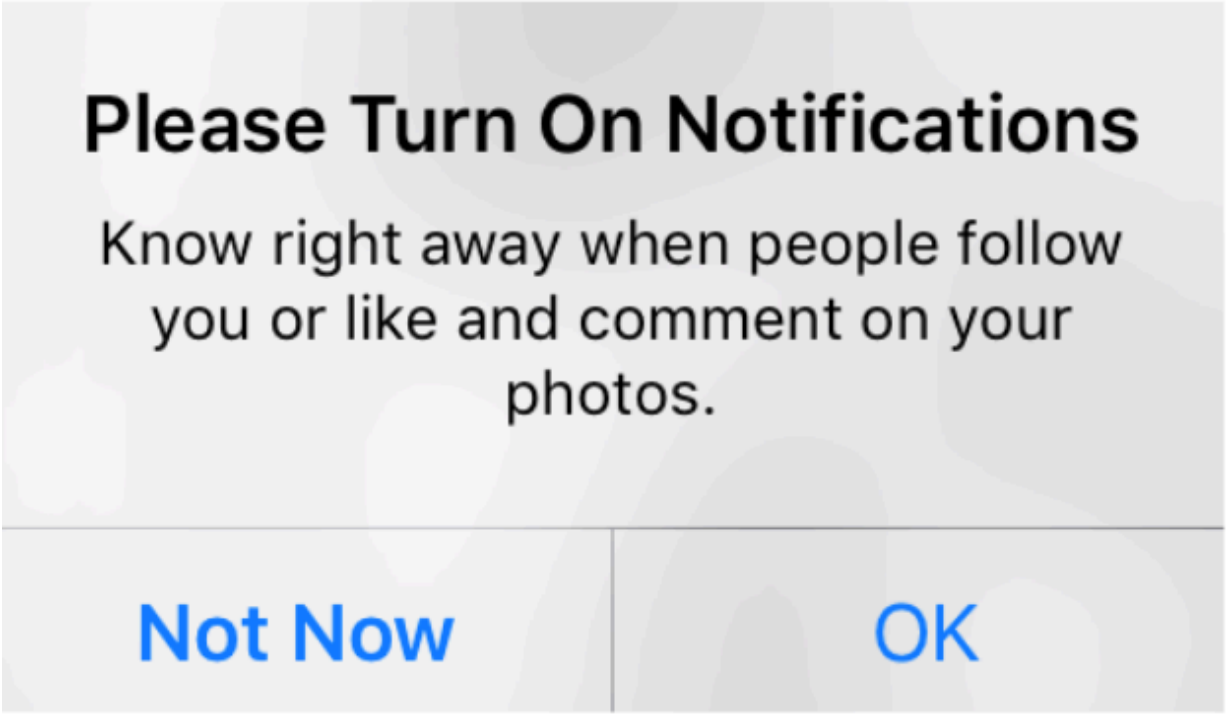
Bait and Switch is when a design communicates that a certain action will lead to a specific result, but in reality it does not.[9] For example, Facebook may notify their users that there is a message for them in their inbox on Facebook. After opening Facebook in their browser or their app, users find that there is not in fact a message there. Instead, that message alert is primarily a ploy to get users to log in to the platform.



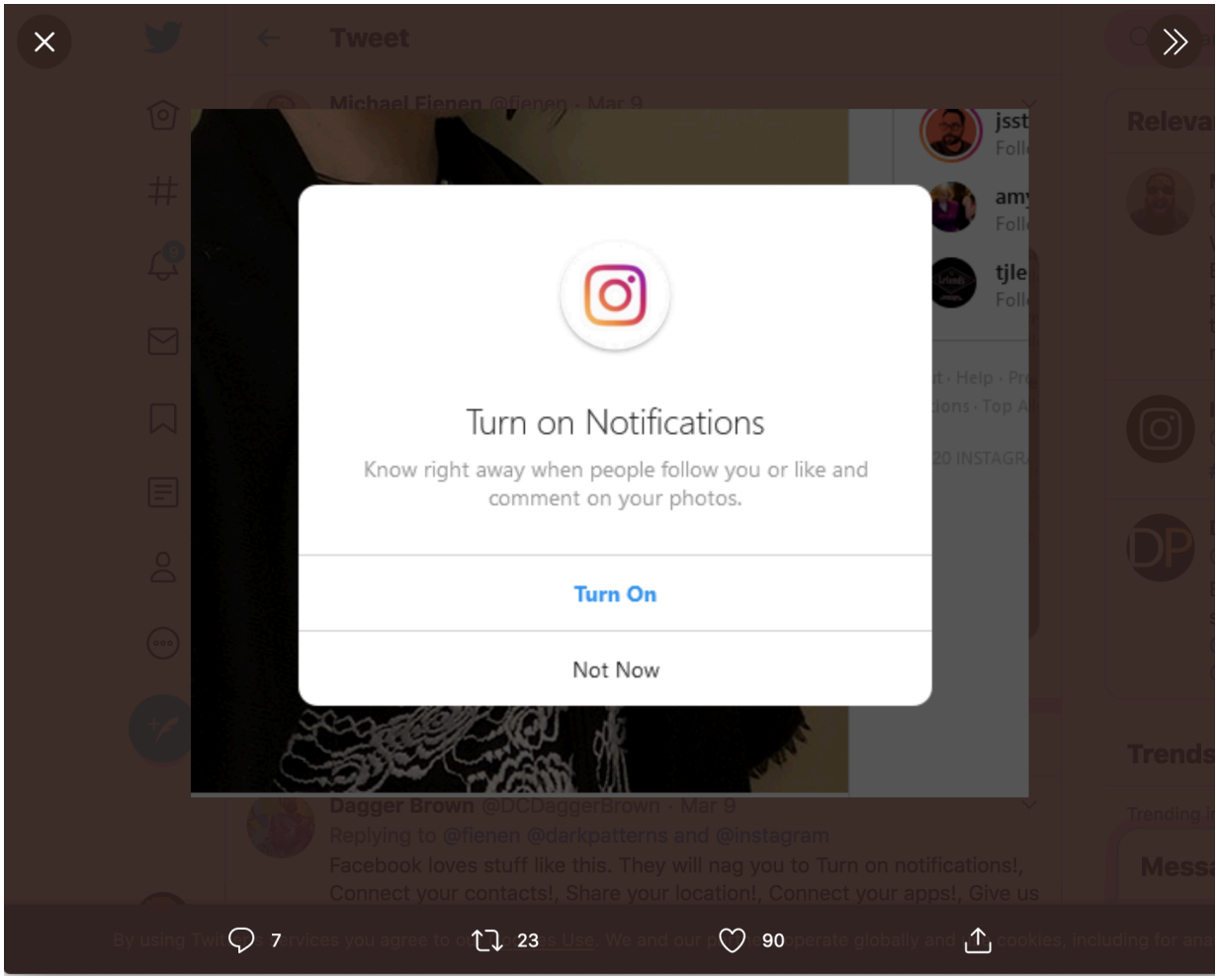
9 Visit: Darkpatterns.org

1.1.5 Nagging

With nagging, the platform provides incessant prompts in an attempt to direct users towards a particular action.[10] One example of this is Instagram's persistent request to users to turn on their notifications to alert them when their post has been liked, or a comment has been made on your post. With this particular prompt, Instagram hopes to capture the interest of the user and translate it into time-spent on the platform. Note the opportunity is not provided to permanently dismiss the message. Instead, the user can choose between temporarily dismissing it (so that it can reappear later) or accepting it. One could argue this does not offer agency, but rather tests users' perseverance: is their desire to resist this particular function strong enough to withstand the annoyance of repeated prompts?[11]



10 "The Dark (Patterns) Side of UX Design".
11 Visit: Darkpatterns.org



1.1.6 Obstruction

Obstruction is, making a particular interaction, such as logging off, more difficult than it needs to be, with the consequence of dissuading it.[12] A particularly common design technique to carry out this form of manipulation is with the placement and design of menus.[13]

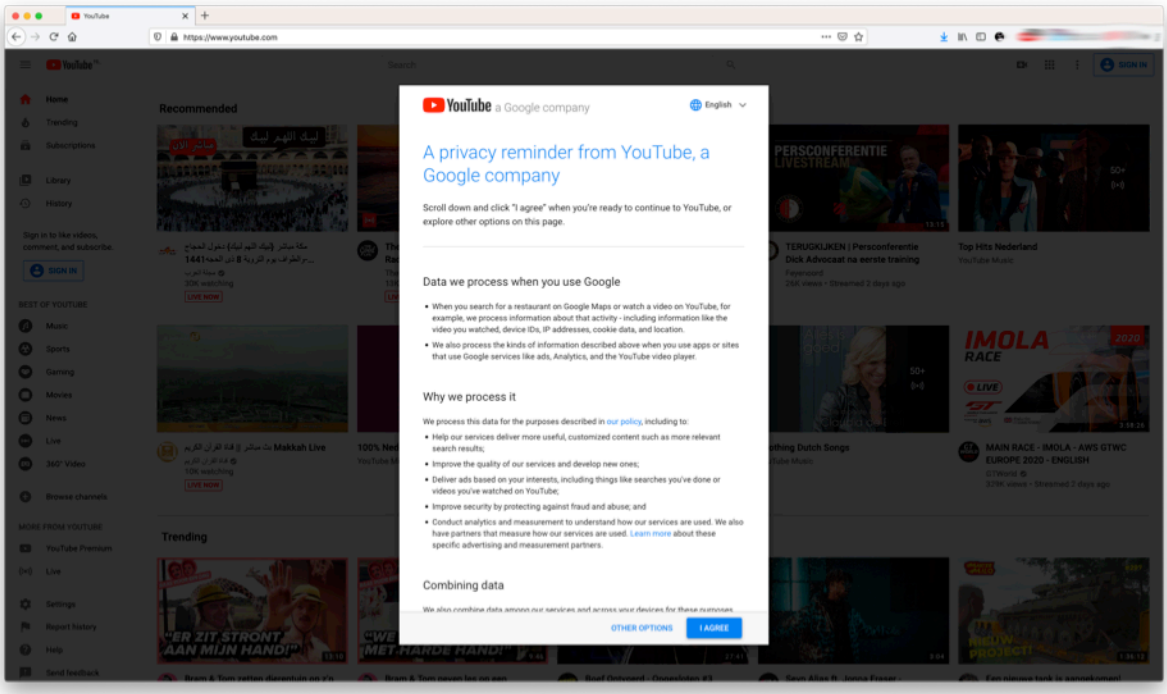
One study on user behavior has found that the design of menus heavily impacts how likely people are to engage with them.[14] For instance, people

12 "The Dark (Patterns) Side of UX Design".

13 Hindering a person's ability to access and change the privacy settings of their account was coined "Privacy Zuckering" in 2010 by Tim Jones. This dark pattern is typically executed through complicated or cumbersome navigation through settings and menus, which are an effort to trick users into accepting the platform's default privacy settings. See: Eff.org and read more about the category on Darkpatterns.org.

14 This study specifically examined GDPR consent menus, however their findings are relevant as they examine how specific design patterns impact user engagement. Source: Nouwens, Midas, et al. "Dark patterns after the GDPR: Scraping consent pop-ups and demonstrating their influence."

tend to engage quite enthusiastically with menus that show all the options on a single page (93.1%), whereas only few users are willing to engage with menus that require navigation to an additional page (6,9%). Menus that feature a toggle button to allow or deny all menu items are much more likely to be engaged with than menus that have a toggle for each individual item. It's not hard to come across examples of platforms utilizing this knowledge to manipulate people into or away from certain behavior. If you want to experience this yourself, browse to the privacy consent menus of YouTube (via Google) or Facebook.



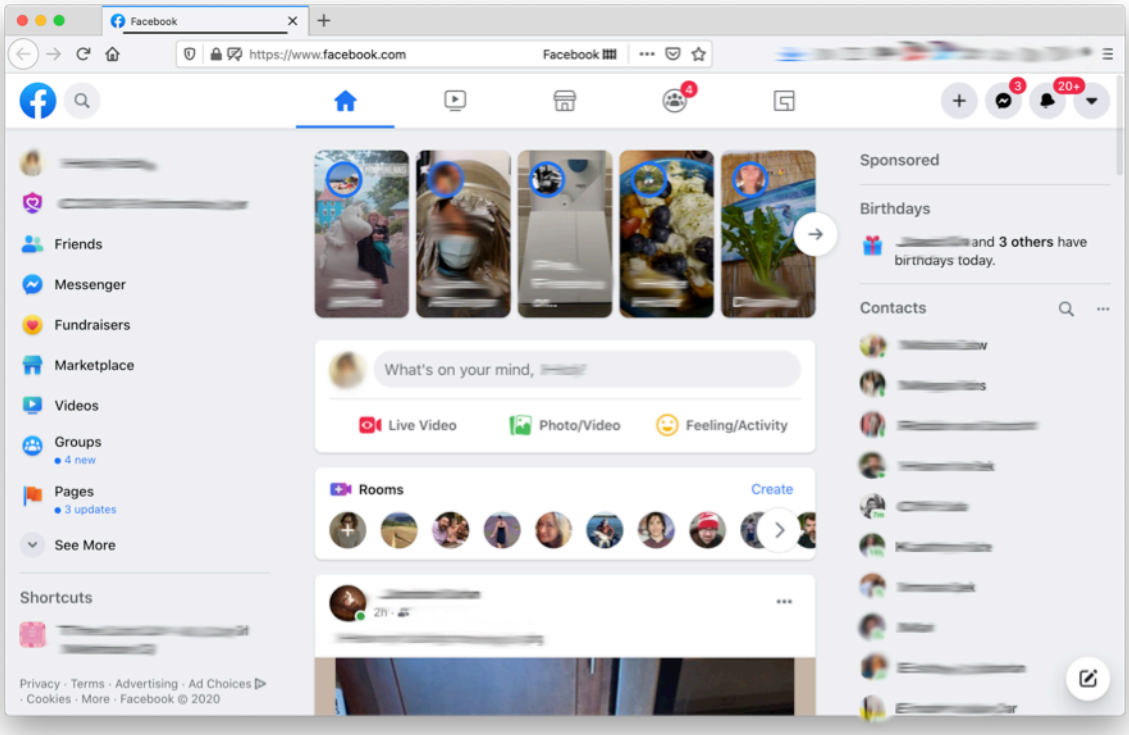
1.2 Prioritizing types of interaction

Not all design tactics related to the layout or interaction design of these platforms are necessarily "dark patterns" that are dedicated to "tricking" users. Other forms of layout and interaction design seek to facilitate or afford particular ways of using the platform; as well as prioritizing particular forms of communication or interaction. These design choices are essential to defining the particular digital service itself.

1.2.1 Facebook

In its current form, Facebook seems to prioritize one-to-one and one-to-many interaction over its typical feed or news content. The layout of Facebook's web-based landing page highlights this agenda. At the top is a short-cut and hook for the stories function. Here, the route for people to engage is via direct and private communication (text or with a pre-

selected set of emojis) and with the original content creator. This landing page is flanked on both sides with shortcuts to community-focused pages on the left, or direct messaging shortcuts on the right. It is only when the user scrolls “below the fold”[15] of the page that the news feed becomes apparent.



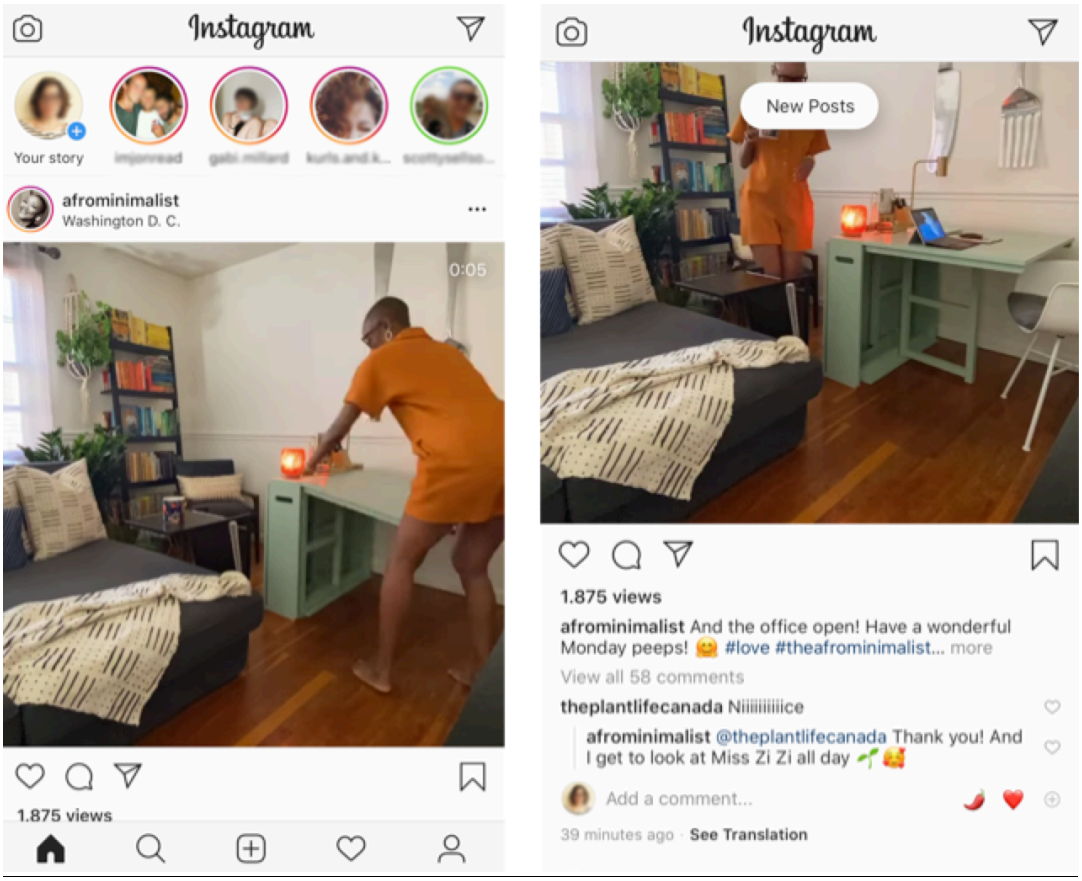
1.2.2 Instagram

Instagram’s core design seems to prioritize exploration and encourages perpetual consumption of content through the delivery of discrete servings via infinite scrolling.[16] Its minimalist design prioritizes images

15 The reference point to being above or below “the fold” harkens back to the large paper format of newspapers. With such large pages, editors selected the most important stories and placed them on the top half of the page. Thus, when the paper was folded (to fit into a bag or a mail slot, or to lay on a stoop) the most important stories would be immediately visible. In the case of digital design, websites or applications follow a similar tradition. Choices about what are the most important things for users to see first are carefully selected. These choices are consistent regardless of how big the screen is that you view it on.

16 Instagram is primarily a mobile app platform, but can also be viewed in a web browser. Even the “desktop” version follows a similar layout design to the app format. Instead of changing the layout of Instagram to take advantage of the real estate available on a computer

that fill the entire screen, reducing distraction. The few menus are embedded in other pages and kept relatively separate from the main feed. At the top of Instagram's landing page, users can find a short-cut to people's temporary content, "stories". Like the infinite scroll, once a user selects one story, another will play immediately after, minimizing barriers to continuous consumption of content. Just below this, and most prominently featured, are users' individual image posts. An individual post has two shortcuts that enable engagement. "Above the fold" (ie, without having to scroll) you can like, comment, share or save; below the fold you can engage with the comment section. These design choices show you are incentivized to directly engage with the original poster over engaging with other commenters.



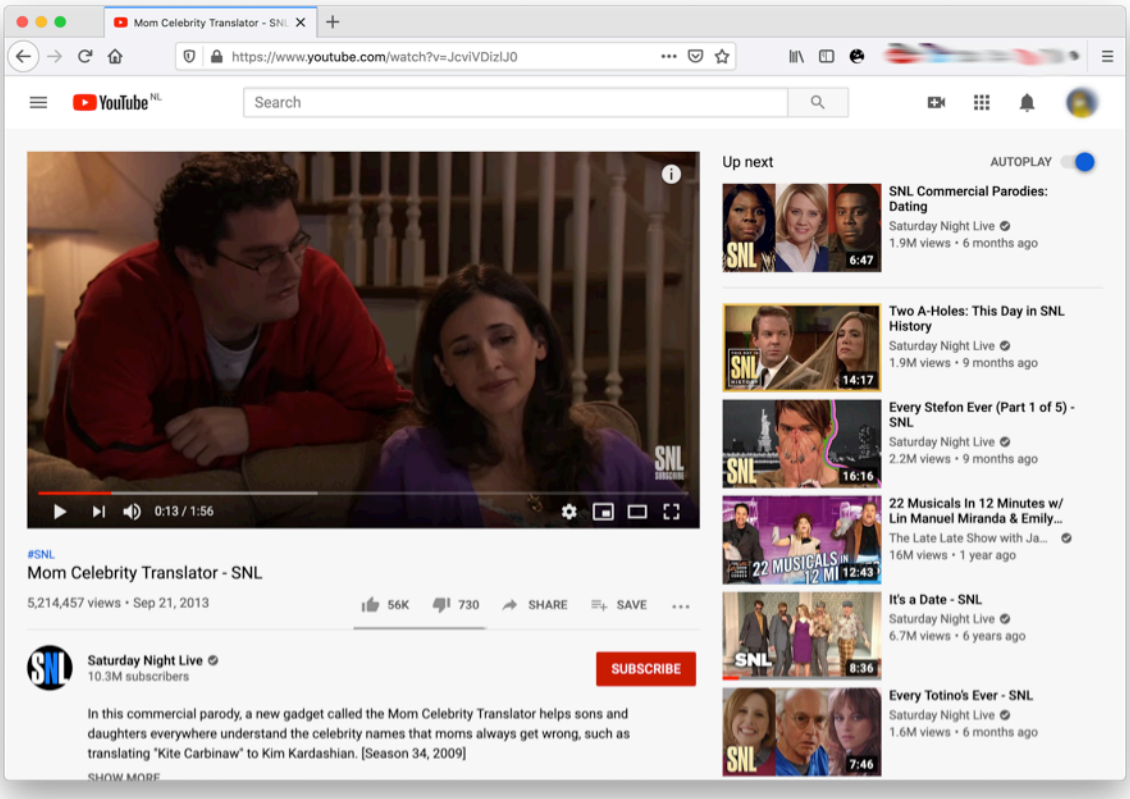
1.2.3 YouTube

YouTube's "landing page" offers a grid of videos for the user to watch, but its recommendation list is the true (and notorious) hook to keep users exploring and engaging with the platform (more on this in chapter 3). After selecting a video, the video consumes about 1/3 of the page.

---

screen, the desktop version of Instagram looks very much like the mobile version.

Exploring is prioritized with the column on the right offering a shortcut to other recommended videos. Directly below the video is a list of engagement metrics such as the number of views, and the number of thumbs ups or down. The opportunity to comment is below the fold.



1.3 On Platform/ Off Platform

Overall, the platforms are designed to keep people engaged on the platform itself rather than being a gateway to other websites. There is, of course, an economic incentive for this decision: more time-spent equals more income. However, each platform offers some opportunities for off-platform traffic. On YouTube, content creators can embed external links in a video caption. These typically become visible to viewers after they have clicked on a (not very prominent) button to expand the caption. Additionally, selected content creators can embed (vetted) external links in videos. On Instagram, too, selected accounts that have reached a particular level of "influence" are granted the ability to embed external links in their stories and posts. Facebook is a slightly different animal. All Facebook users are able to embed a link to content that is hosted outside of Facebook: you simply share a URL. However, in the previous section, we saw that Facebook does not prioritize this type of content and interaction. Instead, it stimulates behavior for which you need not

leave the platform. Facebook's product development reflects this focus. One could argue that the purchase of Instagram and WhatsApp, the creation of Messenger and even that of Free Basics, are ways of keeping users within Facebook's reach and inside the "Facebook family".

How these services appear on your screen is a result of very deliberate choices. Companies are heavily invested in encouraging certain behavior and discouraging others. As this chapter demonstrated, this encouragement is embedded into the choices made about how a platform functions and what it looks like. What is unique about micro-manipulation is that it is perhaps the only form of manipulation that is shown openly, as it interfaces directly with users. This makes this form of manipulation more visible and easier (yet not easy) to critique. For instance, one needs simply to look at a screen to gain a basic insight into if a service's consent interface or privacy settings menu complies with the General Data Protection Regulation. It is interesting to keep this in mind when we explore forms of manipulation that take place in the back ends of these services, and when we dive into systems of manipulation that have a much larger psychosocial dimension.

2. CONTENT MODERATION

Perhaps the most recognizable and widely discussed form of content manipulation is referred to as “content moderation”, the analyzing, labeling and removing (and, essentially, approving) of user-generated content. Content moderation can be done through manual and/or automated means, and is generally based on a company’s terms of service or community guidelines. These documents outline what is and what is not deemed acceptable content and behavior, and legitimize the takedown of a wide range of often poorly defined content. The company may also refer to these documents when penalizing a user.

A lot has been written about the harms of content moderation, from the lack of predictability, to the deplorable working conditions content moderators face. In this report, we will briefly touch on a few of the biggest challenges as they appear to users, and that speak more widely to the threats content moderation poses to our public debate.

2.1 One billion people, one set of rules.

One can imagine that, with a user base as large as that of Facebook, Instagram and YouTube, it is impossible for a single rule to speak across all users, contexts and jurisdictions, and across these contexts always strike the right balance between protection of speech and protection against speech. A clear demonstration of the shortcomings of guidelines and content moderation can be found in the debate about how the depiction of women’s breasts and breast-feeding has been managed by Facebook.[17] Originally banning all female nipples and aureoles because of their “pornographic” nature, Facebook succumbed to public pressure and has been developing, according to the platform itself, a more nuanced view on the female body, eventually giving its thumbs-up to the depiction of female nipples when in the context of active breastfeeding. This has of course raised a lot of new questions. What if a woman’s breast was exposed and a baby was sitting in her lap, but not currently latched? How old can that child be for breastfeeding to still be considered as “appropriate”?

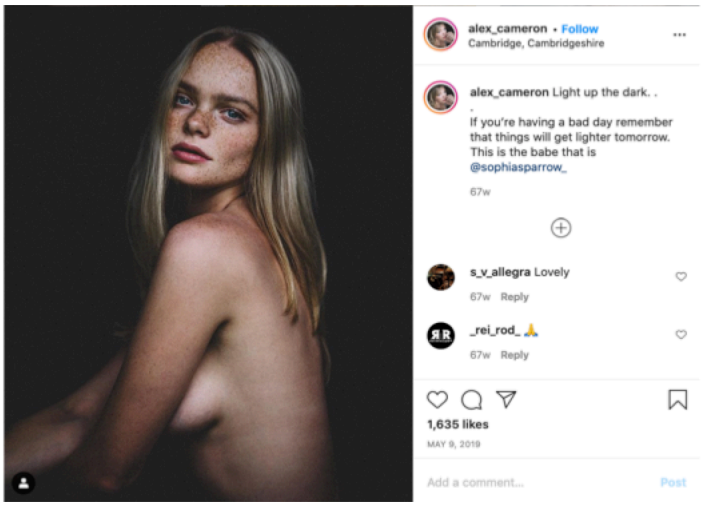
Rules, in other words, are often written with one particular context or use case in mind (pornographic nudity in this case), but fail to account for different cultural and bodily expressions (breastfeeding, and how that may vary across cultural contexts). This example not only highlights the impossibility of catering to such a large audience with one set of rules, it

17 The podcast Radiolab did some great reporting on this, and content moderation more generally.

also illuminates the shortcomings regarding who is envisioned as Facebook users and what they consider appropriate (this is a question of inclusivity), as well as the mechanisms of enforcement. It also illuminates the shortcomings of having a small and rather homogeneous group of people responsible for developing a service that is used by a diverse, global majority. Finally, one must ask themselves what avenues truly exist for society, policy makers, or users to challenge these rules, and what this means for our public debate.

2.2 Lack of predictability

When speech is restricted by law, it needs to happen in a predictable manner. On social media platforms, it largely remains unclear why some content is moderated and removed and other content isn't. A recent controversy on Instagram illustrates this unpredictability. A specific image of a partially nude model was consistently being removed by the platform, while other images of a similar subject (captured by the same photographer and posted by the same account) were not removed.[18] The only apparent difference in these images was the race and body type of the models. The image of the partially nude large black woman was consistently taken down by the platform, while a number of images from the same account of partially nude slim white women remained.



These images, both from the same photographer, both feature women who are nude from the waist up and where outlines of their breasts are apparent. However, it was only the image of the black woman that was repeatedly taken down for violating community guidelines.

18 Nosheel Iqbal. "Instagram 'censorship' of black model's photo reignites claims of race bias". The Guardian.

In the absence of transparency from Instagram, there is only speculation as to why and how one image is blocked and the other is allowed to remain. Is it individual users flagging one piece of content and not the other? Or is it algorithmic bias and does the data set that the algorithm was trained on discriminate against black or fat bodies, or recognize only certain types of images as art or as beauty?[19] Or is there no reason and is it a random occurrence?

2.3 Lack of transparency

Governments, civil society and academics are trying to address the harms to speech caused by platforms, but a lack of transparency makes it impossible to provide meaningful remedy. For instance, Facebook claims that in the first quarter of 2020, 9.6 million pieces of supposed hate speech were removed, up from 5.7 million pieces in the last quarter of 2019. Facebook claims that in the first quarter of 2020 its algorithms detected 88.8% of those posts before users reported them, up from 86% in the previous quarter.[20] Although obviously meant to dazzle, one could argue these numbers aren't very meaningful. How much of this content was actual hate speech? How much was referred to law enforcement? How much was followed up on? How much "hate speech content" was not detected by the system? What kinds of hate speech were identified? Did they receive equal treatment?

In one study, independent researchers identified 300 Facebook posts that included supposed hate speech and found that only about half of these particular posts were removed for violating Facebook's hate speech rules.[21] Furthermore, this particular study identifies an inconsistency in how different types of hate speech are addressed. Cases of racial and ethnic slurs seem to be more rigorously enforced than cases of misogyny.

In the absence of transparency regarding the mechanisms and processes through which these decisions are made, it is not apparent how to address and correct for these issues. If it were a question of algorithmic bias, new data sets could be conscientiously curated to train that algorithm with. Perhaps it is a shortcoming in image recognition technologies, which misidentifies the subject of the image.[22] If this is an

19 In the shadow of the Black Lives Movement, Instagram's CEO Adam Mosseri recently identified algorithmic bias, especially in terms of how it relates to underrepresented groups, as an area for development. See: Instagram.com.

20 Source: Wired.com

21 Source: Wired.com

22 Image recognition technologies has had a well-documented history

issue of user bias, certain protocols could be instituted to weigh and evaluate reported content. These are different problems that would require different types of solutions.

Furthermore, it difficult to understand the benefits and harms of content moderation, and nearly impossible to hold Facebook accountable. All we truly know is that, in response to lawmakers or advertisers requesting them to take action, Facebook is, indeed, doing "something".

2.4 Whose interests are being served?

If content moderation were to function like other rule-based systems, there would be a set of rules for constituents to abide by, and a formal structure to debate the appropriateness of those rules and how those rules are enforced. Over time, and through formal, hopefully equitable, structures, the rules would get amended, contested, or rewritten to adapt and evolve to the context that they exist in.

However, these platforms' commitment and responsibility is not foremost to the user. Content moderation, one could argue, is the removal of user content on platforms not primarily to protect people, but to protect the platform from the diverse forces that influence its business. This is why there is no representative available for the user to reach out to, no hotline to call to challenge a take down. Speech is freely, discreetly, and inconsistently moderated and manipulated for the purposes of the platform itself, be it political or commercial. Two case studies show this quite neatly.

The first was when President Trump incited violence against Black Lives Matter protests (June 2020). In response to Facebook's decision to be one of the few platforms to keep this content online, major companies withdrew their advertisements from the platform citing its lack of diversity and tolerance for hate speech as a part of the recent campaign "Stop Hate for Profit" (June 2020). Within two hours of Unilever pulling its advertisements on the platform,[23] Mark Zuckerburg publicly announced Facebook's new policies against hate speech and misleading information.[24]

---

of being unable to recognize faces of Black people. See for instance: Wired.com

23 It is worth mentioning that most of the companies that pulled their advertisements from Facebook only pulled the advertisements from the US site, but not the international site.

24 Alex Hern. "How hate speech campaigners found Facebook's weak spot". The Guardian.

The second case study is provided by a report speculating that hate speech, especially targeted at women, was not removed from Facebook because it garnered more user engagement:

> "Given the improvements in artificial intelligence and content removal algorithms, Facebook could choose to remove all instances of certain hateful slurs such as "f*g" or "c*nt" when they appear on the platform. The fact that the company does not immediately take this action suggests that they are more than aware of the financial benefits associated with having hate speech, and the users that consume that content, spending more time on the site."[25]

In other words: bolstering engagement metrics trumps protecting people from false or hateful content. Without at least meaningful transparency, it is unclear how users, policy makers, and critics will be able to effect change.

25 Caitlin Ring Carlson and Hayley Rousselle. "Report and repeat: Investigating Facebook's hate speech removal process".

3. ALGORITHMIC CONTENT CURATION

Each of the platforms discussed in this report curate the content that is delivered to their users. In its earliest inception, content appeared on Facebook's and Instagram's feeds chronologically according to when it was uploaded. These "chronologic feeds" were later replaced with "algorithmic feeds," which prioritized content not based on when it was uploaded, but on various criteria determined by the platform. YouTube's recommendation list (or "Up Next" list) similarly is curated by an algorithm that is designed to identify content that the user would likely watch.

The stated intention of algorithmic content curation is to offer a shortcut to the content that is most "relevant" to the user. One could argue this is, at least partly, economically motivated. "Relevance" increases "time spent" on the platform, which is good for that platform's business. The more time a user spends on the platform, the more opportunity that user has to view advertisements and engage with content. This boosts advertising and consolidates platforms' dominance. Additionally, the more time a user spends on the platform, the more opportunity the platform has to collect data on user behavior, another crucial commodity.

3.1 We asked for "meaningful"...

Algorithmic content curation was deployed amid a growing public concern that users were spending too much time on these platforms. Many users of Facebook, Instagram, and YouTube found themselves prone to loosing track of time and becoming utterly engrossed in their feeds. Critics argued that the platforms were being designed to keep users' eyes glued to the screen and that this "attention economy" has harmful impacts on users and society, leading to polarization, addiction, superficiality, political manipulation, and threats to users' mental health. In response, prominent voices advocated we should be striving towards a world where our "time [is] well spent."[26]

Facebook's response in 2018 was to prioritize the content of users' friends and families. In Mark Zuckerburg's announcement[27] about the upcoming changes to Facebook's algorithm he even referred to, or

26 Tristan Harris and the "Time Well Spent" movement, which popularized this particular critique, has since been reorganized as the Center for Humane Technology.

27 Facebook.com

perhaps more accurately he "co-opted"[28], the motivation for the design changes as an effort to promote "time well spent" on these platforms. To spend time more meaningful on the platform, his argument was, it should be time spent engaging with the content of friends and family.

The natural question that then arises, is how do platforms know who a user's friends and family are? How Facebook determines what the "most relevant" posts are to a particular user is unclear. Instagram pulls some data from the user's Facebook accounts to help identify relationships, although it is not clear what that data is specifically. It also uses some signifiers to determine a user's friends and family,[29] such as if it's a person whose posts the user often comments on, or if the user signed up for notifications for another user's posts. With this category of "meaningful content" being largely defined on the basis of engagement and frequency, these accounts could belong to a family member, but also to a brand or an organization. One is left wondering if the category "friends and family" really has anything to do with actual relationships, kinship or bond.

Ironically, although these measures were taken in response to a public movement suggesting that users should take more distance from social media platforms, Instagram confirmed that the introduction of the algorithmic feed increased the time users spend on its platform.[30]

3.2 ...and were given "engaging"

With "meaningful" and "relevant" content defined on the basis of engagement (clicks, likes, number of followers, comments, etc) as opposed to the quality of content itself (admittedly, a subjective concept) or who is making it, certain troubling patterns emerge. For example, Instagram's curation algorithm appears to amplify content that features more skin. A study by Algorithm Watch found that:

> "Posts that contained pictures of women in undergarment[s] or bikini[s] were 54% more likely to appear in the newsfeed of our volunteers. Posts containing pictures of bare chested men were 28% more likely to be shown. By contrast, posts showing pictures of food or landscape were about 60% less likely to be shown in the newsfeed."[31]

As Facebook and Instagram are not transparent about the inner

28 Theverge.com
29 Vox.com
30 Vox.com
31 Algorithmwatch.org

workings of their algorithm, this study can only demonstrate that this trend exists. It is not possible to know why it occurs. However, a 2015 patent[32] authored by Facebook demonstrates that the computer vision technology behind these algorithms have the capability to make an "engagement metric" which is used to determine whether or not to show an image on a user's feed. What is not clear is if this engagement metric is individualized or if this metric is generalizable to all users. This is the difference between:

1. showing a bikini image only to people who have demonstrated in the past that they like bikini images;
2. showing it to the friends of the person wearing the bikini;
3. showing it to a group of people (such as "young males");
4. or even further, if the algorithms assume that bikinis and bare chests are important content for everyone, showing it to every user.

This example emphasizes the lack of transparency as to how these algorithms work, and the motivations behind them. The lack of transparency makes it difficult to assess the risks and identify possible improvements If it is developers making assumptions about what is engaging and what is not, their biases and agendas need to be scrutinized. If "engaging" is based on a personalized metric, one needs to be wary of filter bubbles, and of users becoming sensitized to the content that is made most available to them.

3.3 Facebook and the (de-)prioritization of the news

The impact of this type of curation becomes clear when we look at how Facebook deals with news content. From questions of what types of compensation is owed to publishers,[33] to the virility of misinformation and "fake news" on its platform, Facebook has had to consistently re-evaluate its association with news items.

It goes without question that news items featured on Facebook have a pervasive reach. Additionally, research has found that fake news items spread faster on this platform than any other[34]. The tension between the

32 USPTO Patent Full-Text and Image Database.

33 Australia, for example, is considering a law that would require Facebook to pay publishers to distribute their news content on their platform. To avoid this responsibility, Facebook is threatening to pull all news content from its site in Australia. See: BBC.com.

34 Although, the same report also suggests that the "widespread speculation about the prevalence of exposure to untrustworthy websites has been overstated." See also: Forbes.com.

social media platforms and news content was especially intense surrounding the 2016 US presidential campaign, where Facebook amplified and widely circulated misinformation. Facebook's move towards the algorithmic feed in 2018, supposedly meant to prioritize content from user's network of friends, and de-proritize other sources such as news outlets, actually resulted in higher engagement with news items than in previous years, and more engagement with sensational content.[35] For example, the angry emoji dominated many pages, with Fox News' content earning more than twice as many angry reactions than any other outlet.[36]

Fast forward to June 2020, and Facebook announced it was taking criticism seriously and pivoting away from a news feed that prioritizes engagement, and again had redesigned their algorithm, now to make "original reporting" more visible.[37] In this newest iteration of Facebook's algorithmic feed, Facebook states that the curation algorithm will minimize the noise and amplification of particular content (such as spamming content over distributed networks) and will instead only feature content from its point of origin. Similarly, Facebook says the algorithm will attempt to identify the merits of the new source by promoting content with a byline and publications that have a completed "about" page. Content lacking this information will be de-prioritized in the news feed. Additionally, this iteration of the news feed algorithm includes a labeling system that indicates if a news item is more than 90 days old. Of course, byline and an "about" page for a publication, which the algorithm uses as qualifiers for reliable sources, can easily be fabricated. It is doubtful if the other changes will sufficiently address the previous algorithm's problems and prevent the amplification of misleading information from unreliable sources.

3.4 YouTube's selective success

The recommendation algorithm drives 70% of view-time on YouTube.[38] These algorithms have demonstrated a pattern of promoting sensational or provocative content, as this content typically yields higher user engagement.[39] Facing criticism, in 2019 YouTube took measures to reduce its recommendations of "borderline content that could misinform users in harmful ways — such as videos promoting a phony miracle cure for a serious illness, claiming the earth is flat, or

35 Niemanlab.org
36 Ibid.
37 See Facebook.com and Gizomodo's reporting.
38 Cnet.com
39 See: "A longitudinal analysis of YouTube's promotion of conspiracy videos" by Faddoul, Chaslot and Farid; Nytimes.com.

making blatantly false claims about historical events ”such as the moon landing or of the events of 9/11.”[40] This largely came in the form of limiting access to particular channels.

Research has found that these efforts to curtail the problematic content were initially very successful; however, the results did not persist. This rebound of problematic content is speculated to be due to content creators finding creative ways to work around the constraints of the moderation system; the switch from a manual moderation system to an automatic one; or YouTube relaxing its criteria because of lower engagement or user dissatisfaction.[41] In spite of the initial downward trend, the volume and frequency of conspiratorial content being recommended from information channels remains relatively high.[42]

Interestingly, YouTube appears to have been more successful in limiting misinformation surrounding the Corona virus.[43] In this case YouTube has demonstrated that it has the technology and the means to limit how borderline content is distributed. Critics have argued that effective content curation therefore might be a question of policy, rather than of technology.[44] What needs to be examined is not just the capabilities of these platforms to perform certain tasks or provide certain services, but also how choices are made (or not made) regarding how to exercise and develop their technological capabilities.

40 See: “A longitudinal analysis of YouTube’s promotion of conspiracy videos”.
41 Ibid.
42 Ibid.
43 Ibid.
44 Ibid.

4. MICRO-TARGETING

Measuring and quantifying user behavior, both on- and off-platform, and commodifying these insights, is at the core of these platforms' business model. Platforms use racial, economic, ethnic, or other characteristics or behavior patterns to make generalizations about their users as a basis for determining how external parties engage with that person through their platform. For example, Facebook sells access to over 29,000 unique categories of users. These categories can be dizzyingly specific, such as white boat-loving, classical music listening, former soccer playing women. These categories can become so fine-grained that the targeted group can be made as small 10 people.[45] This is also referred to as "micro-targeting" and is done for the benefit of the particular external party, with the user having limited or no knowledge of the profiling, targeting and personalization of content taking place.

4.1 From personalization to discrimination

There is an established record of advertisers choosing to exclude certain demographics in their marketing campaigns. Facebook itself has even facilitated that discrimination by allowing advertisers to exclude people of certain races from their advertising campaigns, such as for housing or mortgages.[46] There is also a record of companies advertising jobs using gender and age as a means to filter for whom certain listings are made visible to.[47]

Between 2016-2018 the American Civil Liberties Union (ACLU) brought five discrimination lawsuits against Facebook for excluding people from seeing certain housing, employment and credit ads based on gender, age and where they lived. Additional class-action law suits have been brought against Facebook as recently as 2019 for not showing financial services ads to women over a certain age.[48] In March of 2019, Facebook announced that advertisers running housing, employment and credit ads will no longer be able to target users based on age, gender or ZIP code, and will have fewer options when it comes to targeting users.

4.2 Political profiling

Micro-targeting is not just manipulating users in the service of

45 See: "Micro-Targeting and ICT media in the Dutch Parliamentary System". Hazenberg, Van den Hoven, Cuningham, Alfano, Asghari, Sullivan, Ebrahimi Ford, Roriquez.

46 Propublica.org.

47 Nytimes.com.

48 Cnet.com.

commercial interests, but also in the service of manipulating political and democratic processes. This is not a new practice. In the 1960's US presidential election a computer program called a "People Machine" was designed to predict and manipulate human behavior.[49] This initiated an era where data was married with behavioral science research to carve out a new obsession with data and prediction. This practice 'climaxed' in 2016, when the Cambridge Analytica-Facebook data scandal was revealed. Cambridge Analytica had collected data on 87 million Facebook users[50] and used it to generate a "psychological profile" of people. This profile was used, in the 2016 US presidential elections as well as in the run-up to the 2016 Brexit vote, to determine what types of curated political messaging or advertising would be most effective for that particular individual. For example, in the case of Cambridge Analytica's involvement in the Brexit election, it was discovered that certain users identified as "persuadable" were targeted with political advertisements evoking fear regarding the impact to Great Britain when Turkey joins the European Union. These advertisements were targeted towards communities that were economically struggling, such as former coal mining cities and villages. There was, of course, no truthfulness to the claim that Turkey would be joining the European Union, and most would consider targeting communities with struggling economies with a fictitious threat of a sharp sudden influx of competitive laborers, to be exploitative.

This was not the only occurrence of micro targeting being used for the distribution of political messaging. The Global Disinformation Order study,[51] conducted by the University of Oxford, found evidence of social media manipulation by a government agency or political party in 70 countries, an increase from 48 in 2018 and 28 in 2017.[52] The tactics used in these campaigns run the gambit, 75% of which involved circulating seems, fake news, and videos. However more covert methods were also used, such as sponsored trolls to attack opponents such as journalists and activists, tools to censor speech, or promoting particular hashtags to ensure the spread of particular messages.[53] While these tactics are primarily deployed on Facebook, there has been an increase of these types of campaigns focusing circulating photos and videos on Instagram and YouTube.

---

49 NPR.com and *If Then: How the Simulmatics Corporation Invented the Future* by JilL Jepore.

50 Only 270,000 users had directly consented to Cambridge Analytica's data collection. See: Wired.com.

51 See: *The Global Disinformation Order 2019*. Bradshaw, Howard.

52 Digitaltrends.com

53 Ibid.

4.3 What's new?

4.3.1 Specificity

Although not a new practice, the sheer volume of data available (collected both on- and off-platform), and the micro-targeting services offered by the platforms, makes it possible to pinpoint, harness, and exploit specific fears, rhetoric, or arguments to affect or manipulate an individual's perspectives in ways we have not seen before. Compare these mechanisms of profiling to other content and communication "platforms" such as television, print journalism or outdoor billboards. Some demographic profiling occurs here, too. The readers of different newspapers represent different audiences, and advertisers utilize their knowledge of these demographics to target them accordingly. However, these generalizable demographics are nowhere near as nuanced and precise as those that are made possible through the individualized profiles of users on Facebook, Instagram or YouTube.

4.3.2 Openness

Since micro-targeting and personalized content lives in the private screens of individuals and not in the public sphere, there are few opportunities to fully account for these targeted manipulative tactics, or to debate or regulate them. Political actors have no access to the specific debates and claims being made, and thus no opportunity to contest or engage with these claims in public. Micro-targeting might also mislead citizens as to what the priorities of political actors truly are.[54]

4.3.2 Predictability and choice

Further, in most arenas except social media platforms, the audience is self-selecting, and this comes with a certain degree of predictability about how you might be profiled. Consumers can pick and choose which newspaper, shop or magazine they feel most comfortable with. They can discern, based on what they know about the company in question and the "product" it has on offer, how they might be profiled. They can exert their consumer power by choosing a different publication if they do not like its advertisers or how those advertisers target them. In the case of Facebook, Instagram and YouTube, users are not in the opportunity to pick and choose the platform that is most aligned with their affiliations: there are no alternatives to turn to.

54 See: "Online Political Microtargeting: Promises and Threats for Democracy". Zuiderveen Borgesius, Möller, Kruikemeier, Fathaigh, Irion, Dobber, Bodo, De Vreese.

5. PSYCHOSOCIAL MANIPULATION

In chapter 3 we discussed ways in which platforms manipulate what content is and isn't shown in order to keep people engaged. Another mechanism to keep users engaged is the deployment of psychosocial forms of manipulation. This comes in the form of endorsements and validations from other users and the platform's strategic notification of these endorsements. The opportunities that these platforms carve out for endorsement or validation from peers encourage users to represent their lives on these platforms in somewhat misleading ways that negatively impacts other users psychologically. And, in a wicked twist, even the positive experiences of connecting with your friends on these platforms can lead users to harmful behaviors.

5.1 Doing it for the likes

Users tend to evaluate themselves and their content based on what has been validated, or "liked," by others. Social reward systems such as the "like" function on Instagram tap into a neurological function in our brains that desires more of these rewards. The design of Instagram for example capitalizes on validation and reward seeking functions and harnesses them to drive users to spend more time on the platform. Instagram's algorithms will strategically withhold notifying users of when their content has received "likes." Users become disappointed when they see that their content has not been validated by other users, or that it has received fewer responses than expected, only to receive a large bunch later. Dopamine centers had been primed for this negative outcome, and then responds strongly to the sudden influx of social praise.[55] Facebook also uses similar mechanisms of manipulating notifications and rewards on their platforms to incentivize users to open their accounts.[56] This feedback becomes a persuasive force of social influence as well. One study demonstrated that teens were significantly more likely to like a photo if other people had liked it too, while also activating the reward centers in their brains.[57]

This careful manipulation and regulation of dopamine has tremendous impact the body, impacting how cortisol, a hormone that triggers fight-or-flight responses, is released to the brain. This same hormone also triggers anxiety responses.[58] Machine learning makes it easier to identify and target the exact balance in these withholding and reward

55 Simon Parkin. "Has dopamine got us hooked on tech?" The Guardian.
56 Trevor Haynes. "Dopamine, Smartphones & You: A battle for your time". Harvard.edu.
57 Stuart Wolpert. "The teenage brain on social media". UCLA.edu.
58 CBSnews.com.

cycles in order to evoke a particular response from individual users.[59] This technique of strategic withholding and manipulation of dopamine cycles is referred to as the "variable-ratio reward schedule," a concept pioneered by behavioral psychologist B.F Skinner in the 1930s.[60] Dopamine releases are also the basis of nicotine, cocaine, and gambling additions.[61] Casino's are also known for utilizing the variable-ratio reward schedule to keep users hooked.[62]

5.2 So good it hurts

Content created to solicit validation from other users often presents idealized and curated representations of a person's life, one that is aesthetically beautiful and features status symbols. The content that users upload, especially on platforms such as Instagram that are image-rich, tends to represent a version of a user's best or ideal life, not necessarily their lived realities. Representations of the mundane tend not to garner validation, engagement, or reactions from other users.

The consequence of this of course can leave the audience of this content with a distorted view or perspective of reality, leading to detrimental physiological implications. It can lead users to wonder: "why isn't my life as beautiful or exciting or glamorous?" or "why wasn't I invited?" In fact, the Royal Society for Public Health (UK) has ranked Instagram as being the most detrimental social media platform to young people as it was associated with high levels of anxiety, depression, bullying, and FOMO ("fear of missing out").[63]

On the other hand, researchers have also found that browsing Facebook momentarily boosts users' self-esteem. Although this sounds positive, that self-esteem boost ultimately lowers the person's self-control. Researchers found that people who use Facebook more tend to have a higher body-mass index (BMI), increased binge eating, carry more credit card debt, and have lower credit scores.[64] The study concluded that Facebook and other social media platforms can have significant effects

---

59 This is being pioneered by the start-up Dopamine Labs, which largely works on fitness and financial apps. See: The Guardian. https://www.cbsnews.com/news/brain-hacking-tech-insiders-60-minutes/

60 See: "Dopamine, Smartphone & You".

61 Bill Davidow. "Exploiting the Neuroscience of Internet Addiction".

62 *Addiction by Design: Machine Gambling in Las Vegas* by Natasha Dow Schüll.

63 RSPH.org.

64 "Are Close Friends the Enemy?" Wilcox, Stephen. A short video description of the study's findings can be found on YouTube.

on consumer judgment and decision-making.[65]

5.3 Instagram's design response

Responding to criticism, and in an attempt to "depressurize Instagram" (in the words of Instagram executives) Instagram began experimenting with its design to stem the harmful effects of this cycle of validation. The company has been experimenting with removing the "like" feature in an effort to help users become less motivated to compare themselves to others (made possible through the metric of the numbers of "likes"). This experiment has been isolated to users in only a few countries,[66] and the impact of this move is still to be seen. Facebook has also indicated that it's considering removing Like counts from the platform to "present users from destructively comparing themselves to others and possibly feeling inadequate if their posts don't get as many likes."[67]

This is an encouraging step, and it will be interesting to see the effects of this implementation on users. It will also be interesting to see how it will impact the role of engagement metrics in content curation (chapter 3) and how it will impact the economies and markets that have emerged on these platforms, and currently rely on these metrics to determine value.[68]

65 Today.com.

66 Currently: Canada, Ireland, Italy, Japan, Brazil, Australia, and New Zealand. See: PBS.org.

67 Forbes.com.

68 Businessinsider.nl.

6. SELF-MANIPULATION

It is not just a question of the platforms manipulating content and users. Users also manipulate themselves to conform to the expectations of the platform. Self-manipulation is not inherently problematic. In its simplest form, this can be about being respectful to others and endorsing decorum. For instance, it is generally accepted that these platforms should not be a regular home to graphic depictions of abuse or violence.[69] However, at the other end of the spectrum are users manipulating themselves to cater to the goals and agenda of the platform itself, even when these interests are not aligned with nor benefit the user. This is the conceptual core of "manipulation". In these cases, users manipulate themselves and their content to optimize and perform for the platform's key performance indicators: encouraging user engagement and prolonging time-spent. Resources can be found online that "crack" the mysteries of the algorithms to offer aspiring influencers and content creators, or anyone who wants to be more visible online, how to "optimize for the Instagram algorithm": post at regular intervals, post frequently, follow many accounts, use the app frequently and for long periods of time, incorporate IGTV stories (longer viewing period, longer retention and time spent with your content), etc.[70] Visibility can also be bought, such as in buying more followers, buying comments from gig-economy task works, or developing alliances with other content creators to engage with one another's content.[71]

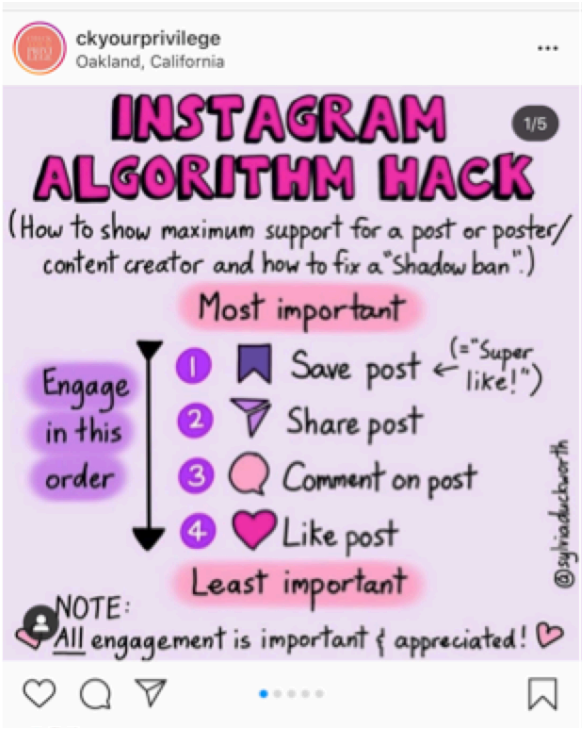
This is essentially an accumulation of the manipulative tactics deployed by the platforms. It is crucial to understand that the platform are designed specifically with this in mind: to incentivize user's self-manipulation to align with the agenda of the platform.

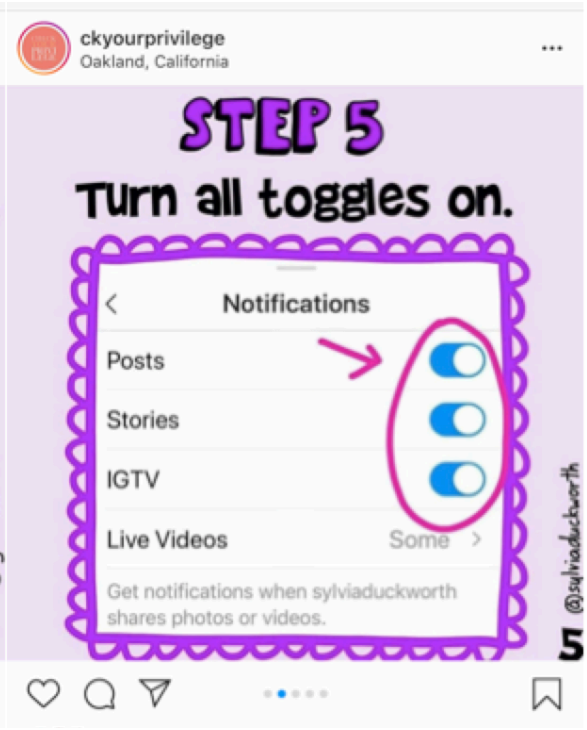
---

69 This is the broadest, most simplistic argument. There are clear occasions when showing images of violence and abuse have societal significance and import. However, these are more often the exception than the rule, and come with careful consideration of senior authorities on the platform.

70 See: Search Engine Watch; Geeks for Geeks; Hootsuite.com; and Hootsuite again.

71 Excellent reporting profiling these strategies and emerging markets dedicated towards promoting visibly on Instagram can be found on VPRO.nl.


ckyourprivilege
Oakland, California
INSTAGRAM ALGORITHM HACK
1/5
(How to show maximum support for a post or poster/ content creator and how to fix a "Shadow ban".)
Most important
Engage in this order
1 Save post ← (="Super like!")
2 Share post
3 Comment on post
4 Like post
Least important
NOTE: All engagement is important & appreciated!
@sylviaduckworth


ckyourprivilege
Oakland, California
STEP 5
Turn all toggles on.
Notifications
Posts
Stories
IGTV
Live Videos
Some
Get notifications when sylviaduckworth shares photos or videos.
@sylviaduckworth
5


ckyourprivilege
Oakland, California
STEP 6
Tap "Live Videos"
Notifications
Posts
Stories
IGTV
Live Videos
Some
Get notifications when sylviaduckworth shares photos or videos.
@sylviaduckworth
6


ckyourprivilege
Oakland, California
THAT'S IT, FOLKS!
Together we will Beat the #$ę@% Instagram algorithm!
@sylviaduckworth
8

6.1 The engagement cycle

As previously discussed, certain types of content tend to get more engagement from users (likes, comments, etc) and are made more visible by the platform's algorithms (chapter 3).[72] For the user, engagement their content translates to validation. Thus there is an incentive to create content that will be rated highly by the platform's algorithms. As chapter 5 explained, the power of these forms of engagement are carefully engineered to feed and regulate the flow of dopamine to users' brains for the purposes of encouraging users to spend more time on the platform, carrying a significant psychosocial and physiological impact.

This creates a cycle in which users create types of content known to be "crowd pleasers". This:

1. Validates the original content creator and keeps them online to receive notifications of how their content is received;
2. Keeps the receiver of this content online and engaged;
3. Contributes to sustaining the economic model of the platform itself;
4. And finally "informs" the platforms' algorithms as to what is engaging content.

6.2 Financial rewards

Platforms also reward and incentivize users with economic opportunities. After reaching a certain level of user engagement, be it follower or subscriber base or numbers of views, users gain the opportunity to earn income off the content they create. YouTube awards opportunities for users to earn income with ads placed in their video content after a certain number of views have been reached. On Instagram, users who reach a certain metric of influence earn the ability to make that influence more efficient and economical. They might, for example, be given additional functionality that makes it easier for them to monetize their Instagram activity. In other words, users are rewarded for contributing to the economy of the platform with the opportunity to take advantage of the platform for their own economic benefit. However, it is important to note that the scale that these economies operate at is vastly unequal, nor is the market fair or regulated.

72 See also: Algorithmwatch.org.

6.3 (Imagined) punishment

There is also a punishment mechanism, referred to as “shadow banning”, for when users do not self-manipulate or comply with the platform’s agendas or biases. This is when content seems to be partially or fully blocked on the platform, but it is not readily apparent if, how or why this is happening. There is no explanation, only the user’s own suspicions of why they may be experiencing a drop in engagement. Did people not like their content; did people not see it; was it (deliberately) buried in the feed; or was there a glitch in the system?

Shadow bans can be especially devastating for users whose businesses depends on the visibility of their content on these platforms; or for users who utilize these platforms for political work or advocacy work. These vested users rely on the platform to function a certain way, however with the ambiguous threat of shadow banning, they become beholden to an agenda or set of guidelines that are not always explicit. As a result, the threat of shadow banning is almost more powerful than the practice itself.

6.4 From self-manipulation to a lack of autonomy

To some extent, it is not uncommon for users to manipulate themselves to behave differently on various outlets. There are different customs as to how to speak and inhabit different spaces. For example, there are different ways in which a person would write an opinion piece for a newspaper, a paper for a school assignment, an email to their boss, verses a message to a friend. In each of these formats, there are power relations to be accounted for, and certain codes and norms and ethical values. Newspapers have editors and journalistic integrity to uphold; teachers and students have rubrics to make evaluations by; bosses and employees have the protocol of contracts, performance reviews, and Human Resources departments; and friends have their own self-determined frameworks of what is acceptable and what is taboo, and how to deal with someone transgressing.

Such a clear set of rules does not exist in the context of these platforms. Furthermore, power is vastly unequal on these platforms, with no way for users to lodge a complaint and expect due process; no way to hold the platform to task. Finally, not all self-expression is treated equal, as some forms come with larger financial benefits for the platform than others. Combined, this comes with great consequences for people’s freedom and self-determination.

7. CONCLUSION

This report offers a taxonomy of six forms of content manipulation that are deployed by YouTube, Facebook, and Instagram. Are the tactics and mechanisms described in this paper in essence a problem? Perhaps not all of them. Design patterns described in chapter 1 can also be a means to help users more readily find their privacy settings. Content moderation (chapter 2) can spare people from encountering deeply troubling content. Algorithmic content curation (chapter 3) can be a mechanism for enabling novel connections between people and between people and information. Micro-targeting can be a way to support local economies or family businesses (chapter 4). Personal bonds can be strengthened with designs that seek to address peer to peer relations (chapter 5). And in theory, some degree of self-manipulation is appropriate when participating in public discourse— one should not yell “fire” in a crowded movie theater (chapter 6).

However, we must conclude that through their market dominance, manipulative practices and lack of transparency, the companies behind the most ubiquitous social media platforms have arguably come to threaten our freedom of expression, self-determination, our public debate and therefore our democracies. This report set out to shed light on some of the mechanisms contributing to these platforms’ dominance. In doing so it wishes to assist both users in better understanding their use of these services, as well as policy advisors and lawmakers striving to mold the digital information ecosystem we need to sustain our democracies.

Bits of Freedom fights for your freedom and privacy on the internet.
These fundamental rights are essential for your development, for technological innovation and for the rule of law. But this freedom isn't self-evident. Your data is being stored and analysed. Your internet traffic is slowed down and blocked.

Bits of Freedom makes sure that your internet is your business.

Bits of Freedom
www.bitsoffreedom.nl
@bitsoffreedom
Prinseneiland 97HS
1013 LN Amsterdam

Contact:
Evelyn Austin
+31 6 2689 5124
evelyn@bitsoffreedom.nl

B5EC 8503 1F6C BEC6 47E6
C0BA E7D0 CB5B 8803 65C9
(bitsoffreedom.nl/openpgp)

BITS OF FREEDOM