Content Moderation Under Pressure: A case study of how Facebook is influenced

Schoof, Celine

CONTENT MODERATION UNDER PRESSURE

A case study of how Facebook is influenced

Celine Schoof

S1688634

Master Crisis and Security Management

Supervisor: Dr. James Shires

Second supervisor: Dr. Els de Busser

04-08-2021

Word count: 16.157

Table of Contents

Introduction

When I was backpacking in South-East Asia in 2019, I have met many people from different countries. Germans, Canadians, Americans, English, French, and other Dutch. To keep in touch, we shared our Facebook or Instagram. Now, I have friends all over the world whom I can still follow and keep in touch with. With billions of users worldwide, connecting, posting, liking, forwarding, replying, tagging, sharing, and commenting 24/7, social media platforms have established an essential role in people's daily lives (Gillespie 2018b; Klonick, 2018). Especially with the COVID-19 crisis that has been going on, and the lockdown that many countries are in, people are reaching out to each other through social media (Koeze & Popper, 2020).

The advent of the Internet, Web 1.0, and Web 2.0 have completely changed the digital landscape. Since the year 2000, social media platforms have grown tremendously and at a rapid pace (Van Dijck, 2013, p. 3-7). The idea of the open web emphasised and expanded new ways of freedom, expression, and knowledge. It allowed people to interact directly with friends and family, and connect with people all around the world via the internet. Moreover, it evolved to a place where people could express and share their ideas and gather in networked publics. In addition, the news brought by traditional media such as newspapers could now be followed live by people on their screens. Nowadays, a handful of social media platforms dominate the social media world. These companies, such as Facebook, Twitter, YouTube, LinkedIn, and Instagram, have millions of users per day (Klonick, 2020; Gill, 2021).

However, there is a downside to this open and free environment with endless opportunities. Since the 2000s, the intended free and open space of social media platforms has also created concerns because it enables the distribution of violent, sexual, illegal, racist, hateful, and fake content (Gillespie, 2018b; Gillespie, 2018a, p. 5-15). Concerns about disinformation, fake news, and extreme violence increased the demand for more regulation of digital platform companies (Gillespie, 2018b; Flew, Martin & Suzor, 2019, p. 33; Napoli, 2019, p. 441).

Social media platforms have implemented guidelines, rules, and regulations that help and prevent negative content mentioned above and help establish a harmonious place for their users. However, the modus operandi, responsibility, and accountability of large social media companies are often questioned by academics and the public. Especially after events that shed a negative light on the way platforms operate. One of the reasons why regulations are being changed is what Ananny and Gillespie (2017) refer to as public shocks. Public shocks are

“Public moments that interrupt the functioning and governance of these ostensibly private platforms, by suddenly highlighting a platform’s infrastructural qualities and call it to account for its public implications.” (Ananny & Gillespie, 2017, p.2-3). These are moments when platforms are called upon to change their functioning and governance. In this paper, I examine three public shocks in detail: Myanmar, Christchurch attacks, and the 2016 elections of the United States of America (US). Specifically, the thesis focuses on public shocks in relation to content moderation policies at social media platforms. Therefore, the research question reads: *How do moments of public shock shape the content moderation policies of social media?*

This thesis aims to provide a more in-depth and clearer understanding of the incentives that social media platforms have to engage in content moderation. This will be examined through the concept of public shocks. The case study focuses on the frontrunner of all social media companies: Facebook. On this platform, I will investigate the three public shock cases mentioned above. Data was gathered through the analysis of primary and secondary sources. The methodology of the case study is further described in section 3 'Methodology'.

Relevance

This research is academically relevant because there has been little case study research on social media content moderation policies concerning moments of public shocks. This study attempts to explain the incentives of social media platforms to engage in content moderation. Previous studies have focused more on research about how social media should be regulated, and about the implications of future regulation. This research aims to contribute to this gap in the literature.

This research is socially relevant because social media plays a significant role in peoples’ daily lives, and is an increasingly important way of obtaining and sharing news and opinions and getting in contact with people. Meanwhile, social media has established itself as a public good. The increasing power of social media platforms results in having a great impact and influence on society, economy, and politics (Gillespie, 2018a, p. 6). Therefore, research on this topic is critical to help us understand this relatively new integral part of modern society.

Concerning the master Crisis and Security Management, it is relevant because the research covers a part on both crisis and on security. As literature shows, the impact of social media platforms is increasing in society, politics, culture, and the economy (Gillespie, 2018a, p. 6). Therefore, it is important to better understand how these platforms work, how they can be regulated, and what the incentives are that influence regulation. This research covers the crisis element by focusing on the concept of public shocks.

Structure

To answer the research question, the thesis is structured in five sections. After this introduction, section 2 provides an overview of the body of knowledge. It begins with a discussion of the broader academic debate on various topics within social media regulation/content moderation and narrows it down to content moderation policies, and to public shocks. Section 3 elaborates on the methodology used. Section 4 follows the analysis on the three used case studies. Lastly, section 5 answers the research question and concludes.

The body of knowledge

Over the last decade, the topic of content moderation has gained much attention from scholars, the public, media, and governments. The academic field of social media is a relatively new and developing one. The new forms of media communication and information through social media have changed rapidly. It opened up a new online world in which it is unclear what rules social media platforms are subjected to, what their responsibilities are, and what forms of regulations should apply to them (Klonick, 2020). The next section elaborates on the body of knowledge.

Firstly, I will introduce the overlap between social media regulation and content moderation. Secondly, I will elaborate on what scholars say about the incentives for content moderation, and lastly, I will discuss the concept of public shock as an incentive.

Regulation, Social media Regulation, and Content Moderation.

To understand the overlap between social media regulation and content moderation, I will first briefly illustrate where regulation itself comes from. As you will see, social media regulation and content moderation are not two completely distinct things.

Traditionally, regulation as a topic is about how states control certain factors/areas of society. The need for regulations comes from the idea that something should be protected. They are implemented in all kinds of areas, for example, the economy, transportation, public health, and the environment. The reason states regulate certain areas is, for example, because they want to provide fairness, equality, and protection (Beales, Brito, Davis, DeMuth, Devine, Dudley, Mannix & McGinnis, 2017, p. 3-4). In social media, it is no different. Here, governments also try to regulate social media platforms to protect people from certain content.

The first categories regulated were on sexual and violent content. Everyone, states and social media companies, will agree that content such as child pornography should be removed from the platform. However, when it comes to removing content concerning ideologies or political views, it is much more complicated (Gillespie, 2018b; Klonick, 2018; Flew, Martin & Suzor, 2019; Napoli, 2019; Citron & Franks, 2020). Moreover, as social media platforms are relatively new, scholars have seen governments struggle when trying to regulate them. Klonick (2018) and Ananny and Gillespie (2017) argue that it is rather difficult because social media platforms on the one hand are private companies, but on the other hand, they fulfil an essential role in peoples' daily lives. They are more or less a public utility.

In traditional media such as broadcasters and editors, media are obliged to regulations set by the government. One may argue to simply transfer these regulations to social media platforms. However, it is not that easy. Effectively regulating social media platforms seems quite difficult because the two have crucial differences. As Pichard and Pickard (2017), and Flew, Martin and Suzor (2019) mention, social media platforms differ in the services they offer and are more complex than traditional media. For example, the real-time pace of interaction and the volume of content largely differ from the content of editors and broadcasters. Another important difference is that editors and broadcasters decide what content they publish, while social media platforms rely on user-generated content and thus do not control what is posted by their users. Furthermore, before content is posted, editors review if the content is true and harmless. But on social media platforms, no one is actively checking content before being posted (Pichard & Pickard, 2017; Klonick, 2018, p. 1660).

If social media platforms cannot be categorized as traditional media, how do states then regulate them? This can be done in different ways and depends on the country. In countries such as Russia and China, social media are more, or fully, controlled and regulated by the state (Lewis, 2017). In Europe, states and the European Union (EU), try to regulate social media by implementing new laws. Social media has to conform to the national legislation, or that of the European Union (Coldewey, 2020). Nonetheless, states are in constant discussion on what rules should apply and how far regulation should go to. On the one hand, they want to protect users, on the other, they do not want to regulate too much so that the freedom, which is so important for social media platforms, is too limited.

In America, regulation is differently arranged. Here, social media enjoy great freedom because they are considered private business companies. This results in not having a legal authority that sets the rules of social media (Coldewey, 2020). Moreover, social media companies are subjected to Section 230. This legislation holds an amendment entitled "Good Samaritan blocking and filtering of offensive content" (Citron & Franks, 2020, p. 4-5). The amendment gives platforms immunity for being held responsible for user-generated content, and at the same time gives them freedom of their content moderation, meaning that they decide what content to keep online and what to delete. This makes Section 230 very valuable to social media platforms such as Facebook and Twitter. Thus, § 230 CDA encourages social media platforms to moderate content themselves. It must be mentioned that when § 230 CDA was enacted, lawmakers had no idea of how the internet would look like in 2021. The legislation is not specifically made for social media platforms, though, social media platforms benefit from it and can operate in a safe harbor (Gillespie, 2018b, p. 203-206; Citron & Franks, 2020, p. 5).

However, it is not only the government that tries to regulate social media companies. In fact, social media platforms are self-regulating. They also want to create a nice environment for their users (Klonick, 2018). One of the ways in which platforms regulate is by content moderation. Content moderation refers to reviewing, reporting, filtering, and deleting content by platforms themselves on their platforms. Content that is being deleted is, for example, violent, racist, and pornographic content (Gillespie, 2018b; Klonick, 2020). Thus, similar to states regulations, content moderation is about setting the boundaries of excepted behaviour on social media platforms. They do that by integrating standards of excepted behaviour. Most platforms have implemented Community Standards, or something similar to that. Along the line of community standards platforms argue to handle content moderation themselves, which does not require state intervention (Klonick, 2020, p. 2428-2429).

The forms of content moderation consist of various ways. The most far-reaching way of content moderation is banning and deleting users' accounts so that they cannot post content anymore. But there are also other ways to moderate content. It can be done by blocking accounts, flagging, warnings, or through Artificial Intelligence (AI) (De Streel, Defreyne, Jacquemin, Ledger &

Michel, 2020, p. 10; Klonick, 2020). Content moderation is important because through this, the platforms protect users and prevents illegal and harmful content to be distributed.

The section above makes it clear that social media platforms are free to decide what content to keep online or to delete. The question then remains, what motives and incentives do platforms have to engage in content moderation?

Incentives content moderation

As private platforms, social media companies are not obliged to have content moderation policies or conform with existing rules (Klonick, 2020, p. 2423). Scholars have argued about what influences the development of platforms content moderation policies. Based on the literature reviewed, this thesis will elaborate on two different academic views.

1. Content moderation is initiated by external factors;
2. Content moderation is initiated by internal factors.

Content moderation by external factors

Scholars have argued that external factors can lead to content moderation change at social media companies. The incentive for change come from external pressure, such as from states, governments, United Nations (UN), European Union, civil society groups, researchers, and media. This is succeeding well. Many content moderation decisions made by platforms are, according to Klonick (2018, 2020), Lessig (2009), Citron and Franks (2020), and Barret et al. (2019), a response to external pressure from one, or more of these entities.

Lessig (2009, p. 19) and Klonick (2018) argue that states try to influence how social media companies operate because they want the companies to comply with existing local laws and legislation. This can be done in different ways. State and governments can do this in a direct way or through an indirect way by lobbying and request. The influence of government on social media content moderation has been apparent over the last few years. The United States, the United Nations, and the European Union requested platforms to moderate content. The video of the Innocence of Muslims perfectly illustrates that platforms indeed conform their moderation following government requests. Platforms decided, after urgent requested by states in the Middle East, to remove the video from their platforms. As a reaction, platforms deleted/banned the video either regionally or globally from their platforms (Klonick, 2018, 1650-1651). This situation shows that platforms 'listen' to states if it comes to content moderation.

A more recent situation in which it is clear that states influence platforms content moderation, is with content and accounts of terrorists of the Islamic State of Iraq and Syria (ISIS). The studies of Klonick (2018), Citron and Franks (2020), and De Streel et al., (2020), illustrate that the European Commission, the United Nations, and the United States, have put pressure on platforms and motivated them to delete hate speech and terrorist activity. As a result, platforms agreed to take measures in reducing terrorist content. Thus, depending on what states label as 'good' or 'bad' content in their country, platforms block and delete certain content in a state or region. Moreover, the case studies of Barrett et al., (2019) also confirm that pressure from governments and states influences how platforms moderate the content. In their research, they illustrate how Facebook adjusted and changed its policies and procedures on international electoral politics. Especially, after the 2016 United States (US) elections, there has been much criticism about the influence of large social media companies such as Facebook during elections. In particular, there was criticism about the bulk of fake news and misinformation, fake accounts run by state actors, and the misuse of private data for micro-advertising (Allcott & Gentzkow, 2017, p. 212-213; Wong, 2019b).

Content moderation by internal factors

External factors can, of course, influence how platforms moderate. However, like Barret, Bridget, Kreiss, and Daniel (2019, p. 15) argue, pressure from external factors does not only lead to changes in policy and procedures. Platforms also change these because of the perceived desirable social ends, values, expectations, and ideals. The previous section implies that platforms only want to act "well/correctly" when external factors put pressure on them. However, some academics argue that the incentives for content moderation come from internal factors, meaning from within the company (Barret et al., 2019; Klonick, 2020; Gallo & Cho, 2021; Gillespie, 2018b; Mazur& Patakyova, 2019; Gill, 2021). These academics state that the platform’s internal incentive stems from the importance of having an increasing number of users on their platforms. This is immediately related to economic incentives. Thus, the internal incentive is economically driven (Gill, 2021, p. 200).

Platforms moderate because they want to generate new users, and keep new users on their platform. Therefore, the trust of the users is very important. The platform must, at least from the user's perspective, be likeable and genuine (Klonick, 2020, p. 2426-2427). Gallo and Cho (2021) agree with Klonick (2020), users are important because their clicks (through advertisements) are the source of income. Therefore, there is a battle going on between social media platforms to generate users and to tempt them to spend more time on their platform than on other platforms. Users expect to see content that is engaging to them. It is the platform's task to live up to those expectations. An important way for platforms to achieve this is by content moderation. This could be done by, for example, change algorithms so that find users see interesting individual-generated content on their timelines (Mazur & Patakyova, 2019, p. 224). In line with this, it is therefore very important for platforms to apply rules on how they want their users to behave on their platforms. Without any rules, there would be an environment where there is anarchy, which most likely results in constant chaos and conflict. No one would like to be there. The internet would then quickly become a place where users would constantly face harmful, violent, hateful, abusive, and illegal content on their timelines (Gillespie 2018b, p. 202). The task for platforms is then, to apply rules to protect one user from another (Gillespie, 2018a, p. 5). Otherwise, a platform would lose its users as they would probably switch to another platform that offers similar products but has a nicer environment.

However, there is a line in the literature saying that divisive content is more engaging. Citron and Franks (2020) argue that looking from the economic drive of platforms (online advertising business model), it is difficult to explain why platforms engage in content moderation. Divisive content grabs more attention and generates more clicks and thus more money. If market forces are most important for social media platforms, it makes more sense to keep divisive content online. Nonetheless, empirical evidence shows that platforms do remove divisive content, and engage in content moderation.

De Streel et al., (2021) shows empirical evidence of internal incentives. In their research, they interviewed employees of several social media companies. Among other things, the interview revealed that social media platforms realized their responsibility as their services took on an increasingly global reach. This is shown by platforms creating Terms of Service/Terms of Use or Community Standards/Guidelines on their platforms that are more or less serve as a 'legal framework'. The legal framework for platform's content moderation is limited. Therefore, their decisions to remove, filter, and block content are based on their Terms of Service/Terms of Use or Community Standards/Guidelines. In addition, the corporate sense of responsibility also includes the responsibility that employees feel. When they face dissatisfaction about how the company handles content moderation, and they express it, it can lead to change as well. This shows that the incentive to engage in content moderation thus stems out of internal factors, and not from states/governments. Quite recently the voluntary contribution has also been made clear through the initiative of Facebook for setting up an independent oversight board for content moderation (Mazur & Patakyova, 2019, p. 223).

As described in the section above, academics have two different views on what drives platforms to content moderation. Some argue that these come from external factors such as governments, political actors, academics, media and civil society, others argue that it is because of internal factors such as economic reasons, corporate responsibility, and employee's dissatisfaction. However, few case studies have been done on the incentives of platforms to implement content moderation policies concerning public shocks.

Public Shocks

Unexpected events with major consequences and high impact can cause a change in society, politics, and people's perceptions. It is therefore interesting to study how these unexpected events can influence a company. Other terms that are used for unexpected events are: crisis, scandals, disasters, etcetera. Concerning social media platforms, Ananny and Gillespie (2017), refer to moments of crisis, scandals or disasters as "public shocks". Public shocks are "*...public moments that interrupt the functioning and governance of these ostensibly private platforms, by suddenly highlighting a platform's infrastructural qualities and call it to account for its public implications*." (Ananny & Gillespie, 2017, p.2-3). It refers to a shocking situation/incident in which the public responds. The shocks can consist of two components:

1). Voids the promise made

2). Makes something unacceptable visible about the current operation of the platform

(Ananny & Gillespie, 2017, p. 6).

Public shocks can result in different cases relating to social media and content moderation. For example, a shock can be caused by content that appears on the platform. It can involve one piece of content or a whole genre. Moreover, shocks can also relate to the way platforms work that is unfamiliar to users and thus come as a surprise. Or it highlights the lack of accountability. Furthermore, a public shock is an incident that is being followed by a public outcry. It can be picked up by the media, the government, and/or by society. Such a public shock always causes criticism, which then results in the public wanting to change platforms, in either specific or general terms. Indignation is a very important feature of public shocks. Therefore, Ananny and Gillespie focus on this factor in their study (Ananny & Gillespie, 2017, p. 6).

However, Ananny and Gillespie (2017) only focus on the indignation that changes platforms operations. Furthermore, they do not see public shocks as a tool that can change platform governance. In their case study they explore the cycle of public shocks in relation to the accountability of social media platforms. According to them, accountability is triggered by a public shock and it leads to unsatisfied and inadequate exceptions on a platform, but rarely to actual regulations. They do not elaborate on external and internal factors as a reason to change regulations. In fact, they do not mention anything about internal factors since public shocks are external itself. On external factors, they argue that it has a minimal impact on change. After platforms react to it (saying sorry and saying that they will change their operations), nothing changes. However, I assume that considering the academic literature, public shocks give reasons for social media platforms to change their content moderation.

To go beyond the study of Ananny and Gillespie, I want to include internal and external factors when examining public shocks as they paid too little attention to internal and external factors. These factors are of influence because, unlike regular private companies, social media platforms have also taken on a public role, and are therefore more responsive to them. In this thesis, unlike Ananny and Gillespie, I want to research the reasons that do or do not lead to content moderation changes. In doing so, I make a distinction between internal factors and external factors. Therefore, I focus on three shocks to research how these influenced social media platform’s content moderations.

A public shock is by definition external. But this does not lead to a change in content moderation all at once. This requires internal pressures or external pressures. Internal pressure come from economic concerns, employee dissatisfaction, or corporate sense of responsibility. External pressure come from the media, governments or political actors such as the EU and the UN, and civil society groups. These factors are the roots that lead to change after a public shock. Moving from these factors, I explore how content moderation change has taken place, and answer my research question.

In the figure below illustrates a schematic representation of the factors that lead to content moderation change.

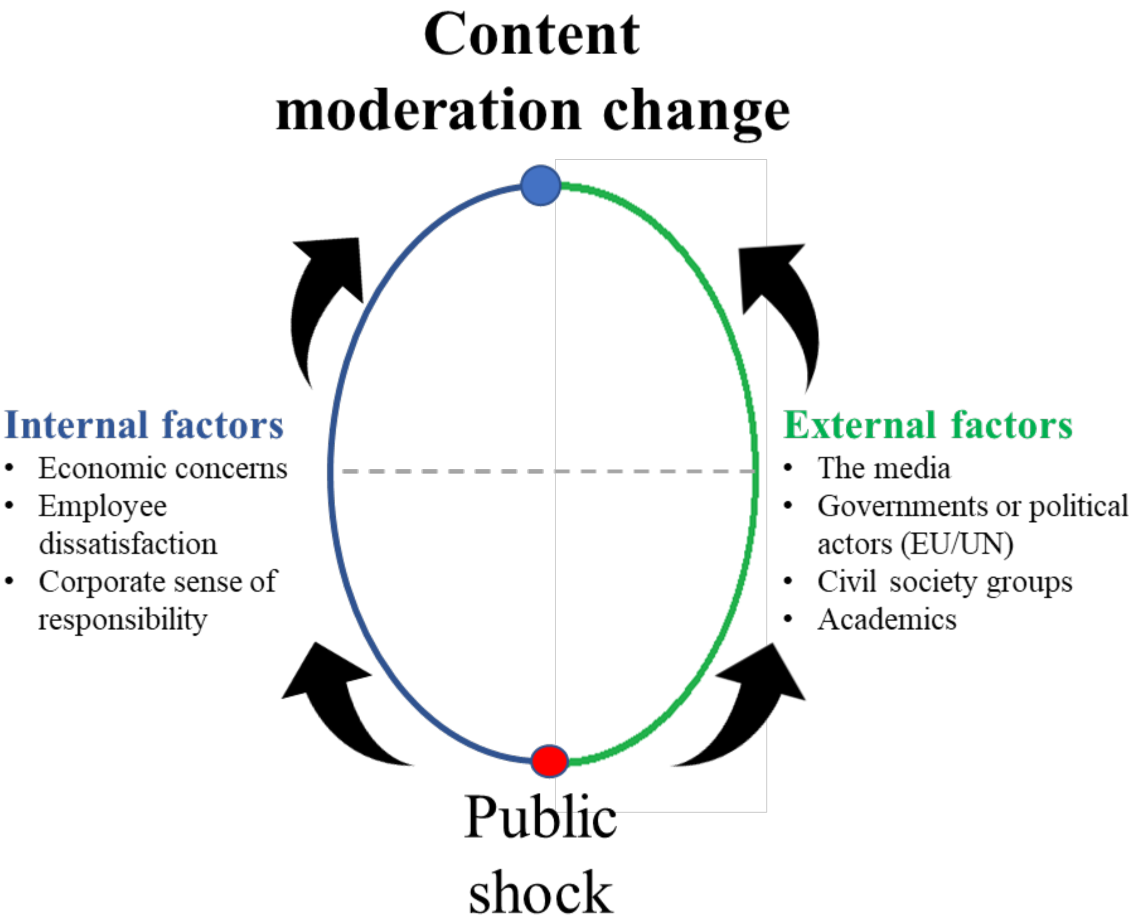
Content moderation change
Internal factors
• Economic concerns
• Employee dissatisfaction
• Corporate sense of responsibility
External factors
• The media
• Governments or political actors (EU/UN)
• Civil society groups
• Academics
Public shock

Figure 1: Factors leading to content moderation change

However, the two factors are not entirely disconnected. Internal pressures can be influenced by external pressures, and vice-versa. For example, in response a media revelation, employees can be dissatisfied and cause internal pressures. This can lead to internal pressure to bring change in the company. Thus, it is not impossible that the two cannot influence each other. Therefore, a horizontal line has been made between the two factors. However, for the purpose of this thesis I distinguish the internal and external pressure each other to give a clearer understanding on how these factors influence change.

Concepts

To answer the research question, I will be using several concepts. Firstly, the concept of social media platforms is used. Nick Srnicek (2017) used the very broad term of social media platforms. He refers to it as a broad spectrum of "digital infrastructures that enable interaction between two or more groups", which consist of "a series of tools that enable users to build their products, services, and marketplaces". This thesis, however, will employ the narrow term of social media platforms defined by Gillespie (2018a) as:

"Online sited and services that:

a) Host organize, and circulate users' shared content or social interactions for them,
b) Without having produced or commissioned (the bulk of) that content,
c) Built on an infrastructure, beneath that circulation of information, for processing data for consumer service, advertising, and profit,
d) Platforms do, and must, moderate the content and activity of users, using some logistics of detection, review, and enforcement."

(Gillespie, 2018a, p.18, p.21).

According to this definition, it includes companies like Facebook, Twitter, and YouTube but excludes marketplaces like Amazon and transportation networks like Uber.

Secondly, the concept used is content moderation. This refers to the set of regulations that social media platforms implemented themselves for reviewing, reporting, filtering, and deleting content from their users (Gillespie, 2018b; Klonick, 2020).

Thirdly, the concept of public shocks by Ananny and Gillespie (2017) is used. Public shocks are "Public moments that interrupt the functioning and governance of these ostensibly private platforms, by suddenly highlighting a platform's infrastructural qualities and call it to account for its public implications." (Ananny & Gillespie, 2017, p.2-3).

Methodology

There is still a lack of understanding about social media platforms and their content moderation policies. Moreover, social media is a relatively new subject, and it changes fast. The section 'Body of Knowledge' illustrates that there are still many question marks around the subject and more research is needed to understand platforms' process of decision making.

To answer the research question, a case-oriented, qualitative, process-tracing methodology centred on Facebook's content moderation is used. Conducting a case study is relevant to collect concrete and in-depth knowledge about a specific topic (Thomas, 2011). This research studies three events, so called public shocks, that happened on Facebook to explore and examine the decisions on content moderation. This methodology is useful because it develops a comprehensive explanation of a process by studying three specific cases in a systematic way (Beach & Pedersen, 2016, p. 306). To understand how Facebook has acted after a public shock, and whether this included internal or external incentives, I combine document analysis, written interviews, official documents, and news articles to understand to what extent public shocks shape Facebook's content moderation.

For this research, I have decided to focus on one social media platform: Facebook. Facebook is a relevant case to focus on because it is, with nearly 2.5 billion monthly users, one of the largest, most widely used, and most well-known platforms (Gillespie, 2018a, p. 20; Gill, 2021, p. 174). Moreover, Facebook has dominated the social media world for almost fifteen years (Klonick, 2020, p. 2418). Because of Facebook's global reach, its impact on society, politics, political economy, and freedom of expression is significant (Maroni, 2019, p. 3). Nowadays, Facebook also owns other platforms that are highly popular such as Instagram and WhatsApp (Shead, 2019). Even though this research will only focus on Facebook itself, it shows the amount of impact that the platform has in the social media world. Facebook is the frontrunner in the social media landscape. Other platforms will most likely also be influenced by how Facebook sets and enforces rules (Gillespie, 2018a, p.20). Focussing on Facebook thus captures a good view of social media platforms.

It must be mentioned that only focusing on Facebook has its limitations. Firstly, it does not include the broad definition of social media platforms. This means, that the research excludes platforms such as Instagram, marketplaces such as eBay, and platforms similar to Uber and Airbnb. Secondly, messaging services such as WhatsApp and WeChat, and online gaming platforms are not researched. Another limitation is that I only do research on Facebook which is highly popular in the West and mainly the United States centred. Other platforms are very popular in other parts of the world, such as VK in Russia, which have a lot of power but are not included in here (Gill, 2021, p. 174). In line with this, I did not analyse platforms that were not in English because of my language skills. However, even though there are limitations, the global reach of Facebook outweighs the positives that my research focuses on.

Facebook has often faced criticism for the way they regulate their content. Some of these criticisms have been widely picked up, investigated, and reported by the media, leading to public shocks. Some events have had more media attention than others. Below, I have made a categorised list of examples of such moments, see *Table 1: Overview of Public Shocks on Facebook*.

To make the categorised list, I used the headings of the Community Standards of Facebook (Community Standards) and added examples of public shocks. The events listed are examples of public shocks. In table 1, I have included a short description on each public shock, and listed under which community standards they fall. The overview of public shocks is in chronological order. For comprehensive information on the community standards, please visit: https://www.facebook.com/communitystandards/introduction.

TABLE 1: Overview of Public Shocks on Facebook			
Year	***Public shock***	***Description***	***Community standards***
2021	• Riot United States Capital	• Former US president Trump made a video and distributed this on Facebook. In the video he encouraged violent behaviour. The video was taken offline too late.	Violence and Criminal Behaviour
2019	• Christchurch Mosque attack live-stream	• Terrorist attack was livestreamed through the 'Facebook-Live' tool.	Violence and Criminal Behaviour
	• COVID-crisis	• The distribution of fake news on COVID-19.	Integrity and Authenticity
2018	• Terrorist activity	• Lack to combat terrorists, violent extremist groups and hate organizations (delete content and accounts). Groups are for example, al-Qaeda, ISIS.	Violence and Criminal Behaviour
	• Brexit	• Fake news on the Brexit	Integrity and Authenticity
2017	• Myanmar	• Thousands of Rohingyas were massacred by Myanmar soldiers. Facebook has contributed to the hate propaganda that has been going on.	Objectionable Content
2016	• Photo "The Terror of War" (Napalm girl)	• Facebook censors the photo "The Terror of War" on their platform because of nudity. However, the photo is iconic for the Vietnam War.	Objectionable Content
	• United States Elections	• Fake news and Cambridge Analytica scandal on data privacy	Integrity and Authenticity
2010	• Breastfeeding photo's	• Photos of mother breastfeeding their babies were deleted by Facebook. Women were angry about the censorship and urged Facebook to change its policy on the topic.	Objectionable Content

Table 1: Overview of Public Shocks on Facebook

Table 1: Overview of Public Shocks on Facebook, contains specific incidents that illustrate different public shocks. Note: this is not a complete list of all the public shocks that happened on Facebook. The list is intended to present an overview of examples of public shocks. Other public shocks that are not listed here may have similarities. Meaning, even though the case was different, the same features were highlighted and criticised. Moreover, in addition to the categories of the Community Standards listed in table 1 (Violence and Criminal Behaviour, Integrity and Authenticity, and Objectionable Content), there are three other subcategories that are also included: Safety, Respecting Intellectual Property, and Content Related Requests. However, since there have not been public shocks in these categories, they are not listed in the table.

The public shocks indicated in table 1 can be divided into the extent that Facebook was involved. Some public shocks were events that did not originally take place on the platform but resulted in public shocks for other reasons. For example, the photo of "The Terror of War". The photo of the Napalm girl was not originally shot and distributed through Facebook. The photo is an iconic photo of the war, and holds a historical message. Though, Facebook decided to censor the photo from its platform. This caused a lot of media attention and criticism. The event itself did not happen on Facebook. However, Facebook was involved because their decision making got questioned by the public and media (Gillespie, 2018a, p. 1-5). Other events make clear that incidents also happen originally on Facebook. These are events where the incident was caused and aggravated by the platform, or where the workings and negative impact of the platform became immediately apparent. The thesis focuses on these events, things that originally happened on Facebook. Even though this is a limitation to the research, these events are more likely to have a greater impact on Facebook.

For the available time I have to conduct this research, I cannot research all the cases. To answer the research question, I will analyse three cases of public shocks.

1. Myanmar
2. Christchurch attacks
3. 2016 US Elections

The cases of public shocks mentioned above are chosen because these events originally happened on Facebook. Moreover, they cover three different elements of the Community Standards, Violence and Criminal Behaviour, Objectionable Content, and Integrity and

Authenticity. In order to avoid being only west oriented, I chose cases that happened in three different continents: Asia, Oceania, and North-America. In doing so, I did not only include developed countries (New Zealand and the US), but also a developing country (Myanmar). The section below elaborates on the cases.

Myanmar

In 2016 and 2017, thousands of Rohingyas were massacred by Myanmar soldiers. This led to more than 800,000 Rohingyas fleeing to neighbour countries (Smith, 2020). The hate for Rohingyas in Myanmar is not something new, however, Facebook has contributed to the hate propaganda that has been going on for quite some years. Most alarming, through this, Facebook has contributed to the ethnic cleansing of Rohingyas. How far the influence of such hate propaganda extends is shown by a UN report of 2018, which called Facebook a "useful tool" for those who want to spread hate (Hoekstra, 2020; Miles, 2018). In Myanmar, Facebook has become more and more important in people daily lives. In 2014, only 1.2 million people used Facebook. That number increased to 18 million users in 2018. Since then, Facebook became the main platform that the people in Myanmar use (Stecklow, 2018). For political parties, powerful generals and extremist Buddhist monks - who label the Rohingya as invaders who want to Islamize the country - Facebook is therefore an important platform with which they can influence their constituency (Hoekstra, 2020). However, hateful content itself is not dangerous. It also depends on the speaker and the audience (Fink, 2018). In this case, hateful content was not shared by regular citizens of Myanmar. Military personnel of Myanmar used Facebook as a tool to spread hate. They did not only do that through their accounts. Posts of them were hidden behind fake names and fake accounts. These stayed undetected (Mozur, 2018). Hate content was then shared up to 9500 times in Myanmar without Facebook taking effective actions (Rajagopalan, Vo & Soe, 2018). Thus, as Facebook did not moderate adequately to counter hate speech, the company was, and still is, being criticised to have contributed to possible genocide.

The case of Myanmar is selected because it took/takes place in a developing country. It is interesting to research whether this public shock affects the content moderation policies of Facebook in the West. Moreover, it is a recent case that has led to the genocide of groups.

Christchurch

On March 15, 2019, a massacre took place at two mosques in Christchurch, New Zealand. 51 people lost their lives that day (Gunia, 2019). What was unique about this terrorist attack was that, for the first time in history, the act was livestreamed via Facebook Live. Facebook Live was used by the perpetrator as a tool to create a video of someone killing Muslims. Through this tool, Facebook users could watch live what the perpetrator did. About 200 people were able to watch the event live and about 4,000 were still able to watch the video before Facebook removed it from its platform. But by then the video had already gone 'viral' and spread through other channels (Macklin, 2019, p. 19-20). This event highlighted the negative side of real-time content and the pace at which dissemination took place. Moreover, Facebook's role in efficiently deleting extreme violent content got questioned by governments and media.

The Christchurch case is chosen because it happened in a developed western country. In addition, it focuses on a tool that Facebook implemented, Facebook Live, but then got heavily misused. Moreover, terrorist content concerning content moderation has been widely discussed in the literature, it is a relevant topic to look into. Also, the Christchurch attack was the first terrorist attack that was being livestreamed (Macklin, 2019, p. 20).

2016 US Elections

Facebook has played an influential role during the 2016 US presidential elections. In 2018 it became clear that, in the run-up to the 2016 elections, data of 87 million Facebook users have been used without their knowledge or consent (Koenis, 2020). Data was gathered by Cambridge Analytica in 2015, which at the time was led by Steve Bannon, who later played a key role in Donald Trump's campaign (NOS, 2018). Through the Facebook-app 'thisisyourdigitallife', data was gathered from users and a database was made. However, people thought their data was being used for academic research, which was not the case. In addition, the data of people from their friend's list was also obtained, unsolicited (Nu.nl, 2018). The data has been used for political targeting to influence voters prior to the 2016 US presidential elections. Through the database, software was being developed to assess Facebook profiles: what will the user vote for? And how can we influence him or her? Personalized political ads could then be created using this information (NOS, 2018). This event highlights the workings of Facebook on data privacy, which caused a lot of criticism.

In addition, there was another incident that led to the 2016 US election being a public shock. In the terms of fake news and misinformation, it appears that Facebook has had great influence in the political sphere. It was not always clear to the user that the news that they saw was fake. According to numbers, fake news articles were shared a lot which increased the dissemination of fake news and misinformation (Allcott & Gentzkow, 2017, p. 212-213). Another revelation involved how Russia had used Facebook as a tool for spreading fake news during the election (Isaac & Wakabayashi, 2017). 470 fake accounts, also called Russian trolls, maintained from Russia got discovered. Through these accounts, at least three thousand political ads were distributed around the time of the election (Koenis, 2020).

The 2016 US Elections are chosen because different revelations about Facebook's workings were highlighted during this event. Moreover, the case took place in the country where Facebook is stationed. Most rules and workings are viewed from an American perspective. The case is one of the recent incidents that has hunted Facebook for a long time, and where CEO Mark Zuckerberg had to testify in Congress regarding Facebook's role in the 2016 US election (Confessore, 2018). Furthermore, this case made countries worldwide question Facebook's influence during their national elections.

To conclude, the three different cases can provide a better overall overview because they took place in different countries, and differ in developing and developed countries, and they cover different elements of the Community Standards of Facebook. Moreover, the three events have led to global (media) attention and questions concerning the way Facebook operates. Furthermore, the three cases mentioned above will help to answer the research question on how moments of public shocks shape the content moderation policies of social media.

Data Collection and Analysis

This research uses qualitative data collection methods. Data will mainly be collected through analysis of primary and secondary sources such as Facebook statements, media reports, indirect information on internal leaked documents, official reports from the governments and political actors such as the UN and the EU, and academic literature. Table 2 'Quantitative indication of sources' provides an overview of used sources.

TABLE 2: QUANTITATIVE INDICATION OF SOURCES	
Myanmar	
Type of sources	*Numbers*
Facebook statements	4
Government/ EU/UN/NGO reports	3
Academic papers	1
Media reports	11
Total	19
Christchurch	
Type of sources	*Numbers*
Facebook statements	7
Government/ EU/UN reports	2
Academic papers	1
Media reports	7
Total	17
2016 US Election	
Type of sources	*Numbers*
Facebook statements	9
Government/ EU/UN reports	1
Academic papers	1
Media reports	18
Total	29

Table 2: Quantitative indication of sources

Selection of cases

There has been extensive media reporting on all three cases. Since I could not include *all* the news reports, I had to make a selection for the analysis. This selection has been done carefully.

In the Myanmar case, I have used the Reuters report and the UN report on Myanmar. These were key reports because exposed the influence of Facebook in Myanmar. These reports led to more criticism. Moreover, The New York Times report by Moose and Mozur (2018), includes the complete and original letter of the civil rights groups, and the response of Zuckerberg. Also, I have included reports in which journalists interviewed employees of Facebook.

In the Christchurch case, I used the information from the Christchurch Call summit because it played a significant role after the public shock. Furthermore, I have used the official report of the Christchurch Call 2021 to examine whether Facebook applied the points they committed to during the summit. The used news reports from VICE written by Cox and/or Koebler, are reports that belong to the subtheme 'content moderation' on its website. These articles provide a broader and more in-depth view about content moderation before and after the Christchurch attacks. Furthermore, they provide more background information on what went on prior to the new policy on white supremacy. In addition, these news reports also include many quotes from Facebook employees.

For the 2016 US election case, I used reports by The Guardian. These were key media reports because they exposed Cambridge Analytica, both in 2015 and 2018. Another news article by The Guardian was used because it shows internal documents from Facebook. These documents contain evidence that Facebook employees were already aware of the concerns at Cambridge Analytica. Moreover, regarding fake news and misinformation, I have used official statements from testimonies or media reports in which statements were quoted. Also, I have used media report of Frenkel (2016) because it includes interviews with Facebook's employees. The report of Fioretti (2018) is chosen because it provides a broader report on EU's perspective, and includes quotes.

Finally, for all three cases, I used Facebook's original statements that they published on their website. I reviewed Facebook's statements made just after a public shock, and the updated statements at a later date to get an idea of how Facebook has reacted and what they have changed. In addition to key reports, I also used many other news reports to get an adequate idea of whether, and what kind, of criticism they have against Facebook. Here I mainly looked at high-quality news websites such as The New York Times, The Guardian, Reuters etcetera.

To analyse the data, the three cases of public shocks will be analysed by researching what incentives influenced Facebook to implement new or modified content moderation policies. To do this structured, the three cases will be tested against four sub-questions.

Sub-questions

1. When did the event happen and when did it become a public shock?
2. What did Facebook do afterwards?
3. What evidence is there of other external pressures? What were these pressures, and what effect did they have?
4. What evidence is there of other internal pressures? What were these pressures, and what effect did they have?

Limitations

Case study

Conducting a case study has its limitations. A case study only focuses on one or a few cases. These cases focus, for example, on the behaviour of a person, a group, or an organization. However, because it is so specific, it is difficult to be certain if the results of one case, also reflect other (similar) cases. This makes it more difficult to generalize the results from case studies. When deciding to generalize it, the findings are more or less suggestive, and additional research is needed to verify the outcomes. Therefore, making causal interferences is difficult and generalization should be done carefully (Simon & Goes, 2013).

Access and selection of (public) data

In this thesis there have been limitations on the access and selection to (public) data. Firstly, media reports provide a huge part of the data for the analysis. However, I could not include *all* the news reports that have been publicized by the media about the cases. Therefore, I had to make a selection which data to use. This results in excluding some information. Furthermore, it should be noted that newspapers are biased because they have their own agendas. They too make a selection in what they write about and in what tone. In addition, I use Facebook's own data which they most of the time publicize on about.facebook.com. Here, Facebook shares statements and offers information about their workings. The statements and information that is written on the website, is written by people working for the company. Therefore, it is important to note that these statements are biased as well. However, as media reports and the reaction of Facebook is very important to my research, I include both of the data in this thesis.

Secondly, another limitation is that I use leaked internal documents from Facebook for my analysis. These internal documents have come into the hands of news agencies. However, they do not publish the full documents. News agencies provide indirect information about what is in these documents. In their articles, they decide which sections to highlight, and share the ones they find most interesting. By not having direct insight into the internal documents, and by the selection that takes place, the credibility of such sources can be questionable. However, since Facebook itself does not make internal documents public, this is the only way to "access" such documents. On balance it is better to use this indirect information, than to not use it at all.

Even though there are limitations to this study, the limitations outweigh the positives that my research focuses on.

Analyses

Myanmar

1. When did the event happen and when did it become a public shock?

Since 2012, Buddhists in Myanmar have been active on Facebook posting hate speech towards (Rohingyas) Muslims. Since then, Myanmar state officials have used Facebook as a tool to post hate content about Muslims or distribute this through their accounts. This has resulted in increased anxiety towards Muslims in the country. Between 2012 and 2014 this led to violent riots between Muslims and anti-Muslims in which hundreds of people died. The public shock was at its peak in 2017 when Myanmar soldiers massacred thousands of Muslim Rohingyas, and some 800.000 people fled the country (Fink, 2018).

2. What did Facebook do afterwards?

In the spring of 2017, Facebook announced that it would take measures to counter hate speech in Myanmar. In 2018, Facebook established new ways of monitoring hate speech in the country. On their website, Facebook stated the following: "Over the course of this year, we have invested heavily in people, technology and partnerships to examine and address the abuse of Facebook in Myanmar, and BSR's report acknowledges that we are now taking the right corrective actions." (Warofka, 2018). Moreover, in an update on the situation in Myanmar, Su (2018) stated that Facebook has enforced its content policies on the following three things:

1. Better tools and technology

This includes investing and extending the use of AI to detect hate speech, accounts that incite hate and violence, and posts that contain graphic violence and comments. In addition to this, Facebook is also hiring more Burmese speaking people that can detect hate speech in Myanmar on its platform.

2. Evolving and enforcing policies

Facebook has updated their credible violence policy. The modified policy now includes fake news and misinformation that may contribute to imminent violence or physical harm. Such content is also removed. Moreover, Facebook is being more pro-active in the detection of hate speech.

3. Partnership and programs on the ground

Facebook is engaging and partnering with civil society to better understand how Facebook's policies are adopted in the country.

Besides that, shortly after the UN criticized Facebook in their report about its contribution to ethnic cleansing in Myanmar, Facebook removed many accounts of Myanmar military officials that incited hate and violence on their accounts (Facebook, 2018f). More recently, in February 2021, Facebook has banned the remaining Myanmar military, and in April 2021, Facebook implemented a new policy to remove praise, support and advocacy of violence by Myanmar security forces and protestors (Frankel, 2021). Even up until now, Facebook is still implementing measures and new policies to counter hate speech.

3. What evidence is there of other external pressures? What were these pressures, and what effect did they have?

As Fink (2018) states, hateful content through Facebook was already posted in 2012. However, Facebook only reacted in 2017. In the report of Reuters, Stecklow (2018) investigates Facebook's failure to effectively respond to the ongoing hate speech in Myanmar. In the report it becomes clear that back in 2013, 2014, and 2015 several researchers, human rights activists, and tech organizations already addressed their concerns to Facebook officials about the spreading of hate speech of Rohingyas on the platform. However, the company did not take any of the concerns seriously back then.

In 2017, the crisis evolved to a critical point when thousands of Rohingyas were mass murdered and some 800.000 people fled the country. Following the UN report on the human rights in Myanmar, it stated: "The role of social media is significant. Facebook has been a useful instrument for those seeking to spread hate, in a context where, for most users, Facebook is the Internet. Although improved in recent months, the response of Facebook has been slow and ineffective..." (United Nations Human Rights Council, 2018). It became clear that Facebook has been a significant contributor to the spread of hate speech and thus fueled ethnic cleansing. Since then, Facebook received widespread international criticism in media reports and by civil society groups.

In a response to Reuters investigation and the UN report, Facebook has acknowledged its slow reaction to counter hate speech in Myanmar (Stecklow, 2018). Sara Su, product manager at Facebook posted the following statement "The ethnic violence in Myanmar is horrific and we have been too slow to prevent misinformation and hate on Facebook." (Su, 2018). Shortly after the publication of the UN report, Facebook announced that they would take the case more seriously and implement new measures (Human Rights Watch, 2019). Facebook's decision to delete the accounts of military and political leaders was influenced by the UN report, as well as by media reports and civil society groups (Slodkowski, 2018).

Furthermore, Facebook started cooperating with civil rights groups, democratic political parties, and the UN (Warofka, 2018; Rajagopalan, Vo & Soe, 2018; Potkin, 2021). In addition, Facebook let an independent commission, Business for Social Responsibility (BSR), assess their role and services in Myanmar (Warofka, 2018). The assessment was not concerned whether Facebook has played a role in Myanmar (the UN report already showed that). The BSR looked at the role and service Facebook has played in the area of human rights. This showed that Facebook's services improves the free speech in the country. However, there are also bad actors taking advantage of the platform by inciting hate and violence on it. To counter latter, Facebook was not adequate in preventing the platform from being used to incite violence. The assessment shows that there is a strong determination inside and outside Facebook to focus on the human rights issue in Myanmar. Furthermore, the BSR came with recommendation for Facebook to improve their role and services in the following five key areas: Governance and Accountability, Enforcement of Content Policies, Engagement, Trust, and Transparency, Advocacy Efforts Aimed at Reform in Myanmar, and Prepare for and Mitigate Risk Related to Future Developments in Myanmar. (BSR, 2018).

Civil society groups also expressed their concerns. In a letter to CEO Mark Zuckerberg, they criticize Facebook's approach to counter hate speech. In their letter, they refer to the inadequate translation systems and the lack of moderators at Facebook who are able to translate the local language, Burmese (Roose & Mozur, 2018). In general, Facebook uses automated systems to detect hate speech. However, these systems struggle to interpret Burmese text because of the way fonts are often displayed on computer screens, making racist and hate speech difficult to recognize (Stecklow, 2018). Because the technology does not work well, Facebook relies on reviewers to analyse the content. However, for years, Facebook did not have Burmese-speaking reviewers in place that could translate the local language. Back then, Burmese content was reviewed by people that spoke English. In 2014, Facebook had hired one Burmese-speaking reviewer, who was stationed in Ireland. At the end of 2015, Facebook had hired three more Burmese-speaking reviewers. These four reviewers had to review content from, the then 7.3 million active users in Myanmar. Moreover, the company did not have an office in Myanmar, therefore, content in being reviewed from abroad (Stecklow, 2018). After the public shock, CEO Mark Zuckerberg told U.S. senators that Facebook would hire dozens of Burmese-speaking moderators (Stecklow, 2018). In 2021, Facebook has over a hundred Burmese-speaking moderators (Perrigo, 2021).

Thus, as the crisis became a public shock in 2017, external factors such as the media, the UN, and civil rights groups have put pressure on Facebook to change its content moderation policies.

4. What evidence is there of other internal pressures? What were these pressures, and what effect did they have?

There is no direct and clear evidence of internal pressures (economic concerns, employee dissatisfaction, or corporate sense of responsibility) before, during, or right after the public shock in 2017 that influenced Facebook's change in content moderation. In fact, after the publication of the UN report on Myanmar, The New York Times interviewed employees of Facebook. According to the interviewees, Facebook is acting successfully when it comes to banning and removing extremist accounts and content that incite violence in Myanmar (Fisher, 2018).

At a later time and indirectly related to the Myanmar case, Facebook employees in general expressed their dissatisfaction and frustrations with Facebook's policies on hate speech, racism and violence. A former employee said that the company has done too little to counter hate speech (Timberg & Dwoskin, 2020). Another employee expressed its frustration to The New York Times stating that the rules on content moderation on hate speech makes him feel like he has killed someone by sometimes not acting (Fisher, 2018).

Another internal pressure that did not follow directly after the public shock in 2017, but does relate to how Facebook combats hate speech, including in Myanmar, is the boycott of several big companies in advertising on Facebook. Companies such as Verizon, Unilever, and Coca-Cola, state that they do not want to have their advertisements running next to posts that incite hate, racism and violence (Reuters Staff, 2020; Paul & Dang, 2020).

The internal pressures described above came from an accumulation of events concerning hate speech. These could have influenced Facebook's content moderation policies on hate speech in Myanmar as well. However, it is not a direct pressure that was caused by the Myanmar public shock in 2017.

Summary

Analysis show that there is no clear evidence on internal pressures that caused change. However, external pressures were most important in this case. Even though the early external pressures in 2013, 2014, and 2015, were ignored, the public shock in 2017 has strengthened these pressures which resulted in content moderation change at Facebook.

After the public shock Facebook has changed and implemented new ways of content moderation on hate speech. Firstly, Facebook enforced its content moderation policies by implementing better tools and technology to hate speech. These changes were only locally implemented to detect hate speech in Myanmar. Next to the technology improvements, Facebook also invested in Burmese-speaking people to review content. In the end of June 2018, Facebook had employed 60 Burmese reviewers. The improvements in AI and investing in Burmese reviewers have contributed to counter hate speech in Myanmar as hundreds of accounts and thousands of posts that involved hate speech were removed. Secondly, the company has changed and implemented its policies on hate speech and violence. Some policies only apply to Myanmar, such as the new policy implemented in April 2021 to remove praise, support and advocacy of violence by Myanmar security forces and protestors. Other policies have been implemented globally, such as the updated credible violence policy, and apply to hate speech in general (Su, 2018; Frankel, 2021). Thirdly, Facebook has made partnerships and programs with civil society groups and local groups. These changes were implemented locally. According to Facebooks update on Myanmar, this has led to important progress (Frankel, 2021).

Even today, Facebook is still trying to improve their content moderation on hate speech in Myanmar as it banned the remaining Myanmar military in 2021 (Frankel, 2021). This highlights that the battle against hate speech in Myanmar is not over yet.

Thus, although external pressures in early reports were not influential, after the public shock of the massacre in 2017, external pressures were most important to influence Facebook’s content moderation policies. The media and the UN were the leading incentives for this. This suggests that public shocks strengthen external pressures to become more influential.

Christchurch

1. When did the event happen and when did it become a public shock?

The Christchurch attacks happened on March 15, 2019 (Gunia, 2019). The incident resulted in an immediate public shock because of the abnormal and horrifying actions. The questions about the workings of Facebook and the tool Facebook Live came shortly after the attack happened. Especially the scale and speed of dissemination of the video was highlighted, and the platform's slow response to remove the video. Furthermore, governments, the media, and civil society groups were critical about the way Facebook, and other big tech companies, countered terrorist content in general.

2. What did Facebook do afterwards?

Right after the attacks, Facebook removed the video, and edits of it, from their platform. Moreover, they deleted the personal accounts of the shooter (Sonderby, 2019). However, this was an immediate reaction that has nothing to do with permanent content moderation policies.

Nonetheless, only one and a half weeks after the attacks, modified content moderation policies were implemented. Facebook strengthened its content moderation policies on the topic of white nationalism and separatism. Content that included this topic is being removed from the platform. This also includes the banning of praising, supporting, and the representation of the topic on its platform (Cox & Koebler, 2019; Macklin, 2019, p. 26).

Furthermore, around May, two months after the attacks and just before the Christchurch Call summit, Facebook tightened their rules regarding Facebook Live. In a statement Guy Rosen, Vice President Integrity at Facebook, announced the 'one strike' policy to Facebook Live: "From now on, anyone who violates our most serious policies will be restricted from using Live for set periods of time – for example 30 days – starting on their first offense" (Rosen, 2019). Furthermore, Rosen announced that Facebook is going to partner with universities and invest $7.5 million in new research for technical innovation to detect manipulated videos and images of the original (violent) content.

During the Christchurch Call summit, Facebook committed to five individual points to tackle violent and extremist content. On their website it says they commit to:

1. Updating the terms of use, community standards, and codes of conduct;
2. Establish methods for reporting terrorist and violent extremist content, through flagging;
3. Invest in technology to improve the detection and removal of terrorist and violent extremist content;
4. Establish checks on livestreaming;
5. Be more transparent regarding the detection and removal of terrorist or violent extremist content.

(Facebook, 2019d).

One year after the summit, Facebook has updated its metrics to disrupt violent and terrorist content behaviour, and increased techniques to detect and delete content related to terrorist groups and organized hate. Facebook's latest update indicates that Facebook, in collaboration with other big social media companies, has created a protocol to "… jointly combat the spread of terrorist content following an attack, established a growing advisory committee of government and international organizations to help inform our work, launched working groups to take new proactive steps to address terrorist and violent extremist content online, and continued to support academic research on how terrorists use digital platforms." (Facebook 2020).

3. What evidence is there of other external pressures? What were these pressures, and what effect did they have?

External pressures came from the media, governments, civil rights groups and academics. Firstly, external pressures came from the media. They have expressed their concerns about how Facebook is being used as a tool to broadcast violent behaviour, and they are sceptic on how the platform is going to prevent this in the future.

Before the livestream of the Christchurch attacks, there were already concerns about the use of Facebook Live. Since the introduction of the tool in 2016, several violent incidents have been livestreamed that included murder and suicide (Seetharaman, 2017). At the time, this also led to criticism by the media about the slow response to remove these videos. In response, Zuckerberg stated to make it easier for users to report such videos, allowing Facebook to respond quicker to remove these videos (Ingram, 2017). There were no additional measures taken at the time.

In 2019, the misuse of the tool was highlighted during the Christchurch attacks. It was the first time that an act of terrorism had been livestreamed through the tool. Therefore, the Christchurch attacks caused an immediate public shock on how Facebook counters terrorist activity.

After the public shock, the same issues were addressed again by the media. This time, media requested Facebook to shut down Facebook Live. Facebook did not comply to this request (Metz & Satariano, 2019). However, it did implement the new restriction of the 'one strike' rule for Facebook Live (Rosen, 2019). In addition, media also criticized the tools and techniques that have failed to detect the Christchurch livestream, and later failed to quickly remove the video. In response to this, and due to the continuous criticism about Facebook's lack of responsiveness to the attack itself, Sandy Sandberg, chief operating officer at Facebook gave a reaction. In a letter she announced that Facebook would explore the implementation of restrictions for using Facebook Live, taking further steps to address hate on the platforms, and supporting the New Zealand community (Sandberg, 2019).

Secondly, civil rights groups and academics have put pressure on Facebook regarding the content moderation policy on white supremacy. Leaked documents show that Facebook makes a distinction between white supremacy, and white nationalism and white separatism. Resulting in only banning content of white supremacy (Cox, 2018; Cox & Koebler, 2018). In 2018, six months before the Christchurch attacks, civil rights groups and academics already expressed their concerns on this topic. According to them, there is no difference in the three concepts. Therefore, they want Facebook to change its policies to ban all three. In conversations with Facebook, they have argued for this. In response, Facebook started reviewing its policy (Cox & Koebler, 2019). However, no explicit changes were made to the policy until after the Christchurch attacks. Two weeks after the public shock, Facebook announced to change its content moderation policy. It now includes the ban of white nationalism and white separatism content. On their website, they stated that the conversations with civil rights groups and academics have convinced them to change it (Facebook, 2019a).

Thirdly, two months after the attacks, governments have addressed their concerns during the Christchurch Call summit. The summit was initiated by New Zealand Prime Minister Jacinda Ardern and French President Emmanuel Macron (Facebook, 2020). During the summit, world leaders and key industry actors came together to discuss how to counter terrorist, and violent extremist content (Christchurch Call).

Originally, Facebook already removed content and accounts related to terrorist behaviour. To do this, Facebook uses AI, human expertise, and partners with other companies, civil society, researchers and governments (Bickert & Fishman, 2017). Through the summit, governments are pressuring tech companies, including Twitter, Google and Facebook, to work more closely together and make profound changes regarding content moderation policies on terrorist and extremist content. In response to the summit, Facebook committed to five individual points to tackle violent and extremist content. Additionally, Facebook acknowledges the influence of external pressures on its website "...The global response to it (the Christchurch attack) in the form of the Christchurch Call to Action, has strongly influenced the recent updates to our policies and their enforcement" (Facebook, 2019c).

The Christchurch Call summit report of 2021 shows that Facebook implemented stricter rules to Facebook Live. Furthermore, Facebook has updated the definition of terrorist or dangerous organizations, and the company now presents a regular transparency report called "Community Standards Enforcement Report". Lastly, Facebook put restrictions on certain hashtags, titles of pages or groups if they are related to dangerous and violent organizations (Christchurch Call, 2021, p. 33-40)

The above shows that there is evidence that external pressures by the media, governments, civil rights groups and academics were present. Moreover, it is clear that after the public shock, partly due to external factors, Facebook improved and changed its rules on content moderation.

4. What evidence is there of other internal pressures? What were these pressures, and what effect did they have?

There is one example of internal pressures that may have led to the change in content moderation. In this case, economic reasons could have led to change in content moderation policies. A few days after the attack, business withdrew their advertisings on Facebook

(Edmunds, 2019). Advertisements are the source of income for Facebook (Klonick, 2019). Therefore, platforms must be credible and trustworthy if they want companies to run their ads on their platform. As a result, this internal driven economic incentive could have influenced the content moderation policies.

There is no evidence found in the data that other internal pressures (employment dissatisfaction, or corporate sense of responsibility) have led to change.

Summary

In the Christchurch case, technological changes were made by Facebook, as well as stricter content moderation policies. Both internal and external pressures are present to have led to these changes. The internal pressure was apparent when major companies withdrew their advertisements on Facebook after the public shock. Because of economic reasons, this could have triggered an internal incentive to change Facebook's policies on terrorist and violent content. However, it is difficult to determine whether this internal pressure alone led to change. In this case, it is more plausible that the external factors, such as media reports, influenced the companies which led to the withdrawal of advertisements. This shows that internal and external factors can reinforce each other and thus are not entirely separate.

Despite the internal pressure, external pressures have been the most important in this case. Firstly, analysis show that the media has influenced Facebook decisions to tighten the rules on Facebook Live. Concerns about Facebook Live were already expressed by the media before the attacks. However, Facebook only made it easier for users to report violent livestreams. The 'one strike' measure taken after the public shock is more profound because, it has to do with the use of the tool itself. Though, Facebook is not doing exactly what the media asks of them (shut down the tool), it does show that the platform is influenced by pressures from the media.

Secondly, after the public shock, Facebook attended the summit together with world leaders and other big tech companies. The summit addressed the issue of terrorist and violent extremist content on social media. After the summit, Facebook agreed to commit to five points concerning this topic. Facebook is still improving techniques and policies resulting from the points of the summit. The Christchurch Call summit report of 2021 confirm that these points have been successfully worked out by Facebook. These changes are implemented worldwide. This shows that Facebook is taking external pressure from the summit seriously and makes permanent and profound changes. The Christchurch Call summit was leading to influence most of Facebooks content moderation policies. This is evident from the fact that Facebook itself has said that the summit has influenced their policy (Facebook, 2019c).

Lastly, Facebook changed its content moderation policy on white supremacy because of the pressure of civil rights groups and academics. Civil rights groups and academics criticized it for a long time, but it was only after the public shock that Facebook changed it. The policy now includes to also remove content regarding white nationalism and white separatism. The reviewed policy on white supremacy is extensive as it is applied globally. Moreover, the change is extensive because it also applies to the popular platform Instagram, which is owned by Facebook. The changes to the policy are permanent. Moreover, Facebook keeps improving their policies, tools and techniques to counter terrorist and violent content.

The analysis shows that, although there was an internal pressure, external pressures were most important. As Facebook itself pointed out, the Christchurch Call summit has influenced, and most certainly fast-forwarded, policy changes on the platform. This suggests that external pressures together with the public shock, influenced Facebooks decision to change its policies.

2016 United States Election

1. When did the event happen and when did it become a public shock?

The 2016 US elections became a public shock for Facebook because of two events. The first event is the Cambridge Analytica data privacy scandal. Cambridge Analytica started in 2015 by collecting data from Facebook users. In December 2015 it was first mentioned in the media when the Guardian revealed an article about the issues of data collection at Cambridge Analytica. However, back then, it did not get much attention yet. The real public shock came later in 2018 when whistle-blower, and former employee of Cambridge Analytica, Christopher Wylie shared his story (Nu.nl, 2018).

The second event is the misinformation and fake news on Facebook during the 2016 US election. Fake news is not something new. Even before the Internet fake news already existed. However, the speed and quantity of dissemination of misinformation and fake news have increased extremely because of social media (Allcot & Gentzkow, 2017, p. 216-217). However, this became a public shock when it became more apparent how, and by whom the US elections were influenced through fake news and misinformation on Facebook. The public shock became even bigger ten months after Election Day, in 2017, when Russia's involvement through fake accounts was revealed (Koenis, 2020).

These two events combined caused the public shock of the 2016 election.

2. What did Facebook do afterwards?

After the public shock, Facebook made several changes regarding data privacy and countering fake news and misinformation.

In response to the Cambridge Analytica scandal, Facebook has made several promises to protect and improve users' data privacy. On their website they stated: "We're going to set a higher standard for how developers build on Facebook, what people should expect from them, and, most importantly, from us." (Facebook, 2018b). To comply with this, Facebook has announced the following changes. Firstly, Facebook has tightened their app control. Facebook hired external partners to make audits of third-party apps. Furthermore, Facebook deleted thousands of third-party apps through the 'App Developer Investigation' that Facebook had initiated (Archibong, 2019; Koenis, 2020). Moreover, Facebook changed its policy regarding who can obtain users' data (Facebook, 2018b). Secondly, Facebook changed its privacy settings. This provides users with more control of their privacy and makes it easier for users to find and change their privacy settings (Egan & Beringer, 2018). Lastly, Facebook announced that they would be more transparent on data privacy. Informing users about what information is collected, and why (Lomas, 2018).

To counter misinformation and fake news, Facebook has focused on the following four main topics since the public shock in 2016.

1. Combating foreign interferences

To combat foreign interferences, Facebook announced a new tool that is more proactive. The tool can detect suspicious accounts such as Russian trolls that post and distribute election-related activity that is fake. These accounts and/or content is then sent to review and can be removed if it is indeed fake.

2. Removing fake accounts

Instead of being dependent on users flagging fake content and accounts, Facebook is being more pro-active in detecting fake news, misinformation, and fake accounts. To make this possible, the company uses improved machine learning tools, and has hired 10,000 people who work in the safety and security domain.

3. Increasing advertisement transparency

Facebook implemented new ways to make advertainments more transparent. This means that users get more insight in advertisements they are seeing, who runs it, and what others ads the advertiser is running. Furthermore, Facebook implemented stricter rules on advertisements, especially those related to elections. People who want to run ads need to verify themselves, and election ads are marked as election ads on the platform.

4. Reducing the spread of fake news

To counter the distribution of fake news, Facebook has deployed new AI tools. Fact-checkers and machine learning tools have fastened the process to detect such news and limit the spreading of it. Moreover, news that is rated as fake, will be labelled so that users will know that they read or share non-verified news. In addition, Facebook also partnered with trustworthy press/news agencies in different countries, such as The Associated Press and the French AFP, to confirm whether news is fake or not.

(Facebook, 2018c)

3. What evidence is there of other external pressures? What were these pressures, and what effect did they have?

Cambridge Analytica

Leading up to the U.S. presidential election, it became clear that the use of data marketing and digital communications were becoming increasingly important during campaign time to target voters. Cambridge Analytica offered Republicans running for office to help them target voters in this way. According to them you can influence people’s personality when you have enough data about them. With this data, you can influence behaviour, and thus the way people vote (Amer & Noujaim, 2019).

Before the public shock in 2018, there were already some external pressures by the media towards Facebook. In 2015, The Guardian revealed that Cambridge Analytica was helping Ted Cruz, Republican presidential candidate, during his campaign. The revelation exposed that the data of Facebook users collected by Cambridge Analytica had been obtained unsolicited, and that the company had violated Facebooks' rules regarding data collection (Davies, 2015). In response, Facebook requested Cambridge Analytica to delete the obtained data but did not take further measures. Due to Facebook's reaction, and other important news around that time, the revelation did not become a public shock. This seemed to have closed the case for Facebook (Thompson & Vogelstein, 2018).

However, in March 2018, the story unfolded into a public shock when The Guardian and The New York Times publicized a series of articles on Cambridge Analytica. Starting with the story of whistle-blower, and former employee of Cambridge Analytica, Chris Wylie. Here, many more details were revealed regarding how the company obtained Facebook users' data, and how they then used that data to influence peoples' political choices. In addition, it became clear that after requesting Cambridge Analytica to delete data in 2015, Facebook had not verified whether this was actually done. Hence, Facebook has been negligent (Cadwalladr & Graham-Harrison, 2018). This report led to widespread outrage over the role of Facebook during the election, and created more external pressures by the media, governments, and the European Union, who called for rapid change in privacy and data protection.

Firstly, the media has put more pressure on Facebook. The story was shared widespread and caused a lot of criticism about how Facebook handles the data of their users. In response, Facebook announced in a statement to suspend Cambridge Analytica from their platform (Grewal, 2018). Facebooks' timing for this was remarkable because it happened a few days before the 2018 revelations were published. However, Facebook knew that The Guardian was working on another story about Cambridge Analytica. After all, the news company had asked Facebook for a comment prior to the publication. It was only then, a few days before the public shock, but two years since the issue was first reported, that Facebook decided to take serious action (Cadwalladr & Graham-Harrison, 2018). From this, it seems that Facebook is influenced by the media in making choices.

The other measures that Facebook adopted after the public shock were both a response to the above-mentioned pressure from the media, and due to political pressures form the US government and the European Union. The seriousness of the situation became clear when CEO Mark Zuckerberg had to testify in US Congress about what happened during the 2016 election (Wichter, 2018). Moreover, Facebook has been feeling pressure from the European Union to change their policy for some time. The General Data Protection Regulation (GDPR), the European legislation on data protection and privacy in the EU, focused on improving the data protection and privacy rules of digital communication (Wolford, n.d). As a result, Facebook is committed to transparency, control, and accountability. In response to the GDPR, Facebook has been working for some time to make its products and services compliant with the rules of GDPR (Facebook, 2018a). Remarkably, after the public shock Facebook has adopted new measures more rapidly. These privacy measures build on the previous measures of the GDPR. In a statement, Facebook acknowledges the influence of the EU is saying "Some of these updates were already in the works, and some are related to new data protection laws coming into effect in the EU. This week's events [revelation of Cambridge Analytica in 2018] have accelerated our efforts, and these changes will be the first of many we plan to roll out to protect people's information and make our platform safer." (Facebook, 2018b).

Fake news and misinformation

In the case of fake news and misinformation, Facebook initially claimed that it was '*crazy*' to think that misinformation on the website had an impact during the 2016 US elections (Levin, 2017), and reluctantly invested in manpower and fighting fake news on the platform (Van Bemmel, 2020). However, the media, governments, and the EU have largely criticised Facebook for not reacting adequately to this issue.

In the media, the topic was abundantly highlighted. Moreover, the media has contributed to the external pressures by publishing the criticisms of governments and the European Union.

Firstly, governments have put pressure on Facebook. The US government has called Facebook to testify in an open hearing before the Committee on Intelligence of the US Senate about the influence of social media in the 2016 US elections. During the open hearing, Facebook spoke about how Russia had misused the platform to interfere in the US elections. In response, Facebook announced several measures to counteract this (Open hearing: Social Media Influence in the 2016 U.S. Elections, 2017). Moreover, CEO Mark Zuckerberg was called to testify in Congress. The topic of fake news and Russia's interference during the 2016 election was also addressed (Watson, 2018). After these pressures, Facebook took several measures. These are mentioned in sub-question 2. Furthermore, legislators from other countries have also put pressure on Facebook. For example, in 2017 Germany has implemented the Network Enforcement Act, NetzDG, aimed at combating fake news. To conform to the new legislation, Facebook implemented a new filtering service in Germany (Bond & Robinson, 2017).

Secondly, the European Union has expressed its concerns. The 2016 US elections made it clear that social media platforms were inadequate in responding to fake news, fake accounts, and the interference of Russia (Bond & Robinson, 2017). This had the EU worried about future European elections. Therefore, the EU requested Facebook to improve its ways to counter fake news and the distribution of it. To avoid further EU measures, they requested Facebook to increase the efforts to remove fake accounts, take action to limit the income of those publishing disinformation, and reduce the targeting options for political advertisers (Fioretti, 2018).

Thus, external pressures were present in both events. In Zuckerberg's testimony to Congress, Zuckerberg acknowledges Facebooks' lack of adequate action saying "It's clear now that we didn't do enough to prevent these tools from being used for harm. That goes for fake news, foreign interference in elections, and hate speech, as well as developers and data privacy.". Referring to the Cambridge Analytica and the fake news and misinformation regarding the 2016 election (Watson, 2018).

4. What evidence is there of other internal pressures? What were these pressures, and what effect did they have?

The two public shocks have changed the way employees view the company. The public shocks have led to a decrease in trust in Facebooks' direction and leadership. In both cases, internal pressure from Facebooks' employees were present.

Cambridge Analytica

Internal documents of Facebook from before the 2015 revelation illustrate that Facebook already knew that there were issues with data collection by Cambridge Analytica. However, this did not relate to the data breach, but to another incident (Wong, 2019a). Still, the documents are valuable because they show how the problems related to Cambridge Analytica were dealt with internally before and after the public shock.

Emails between Facebook employees reveal that there were already concerns about Cambridge Analytica's data collection. These e-mails state: "We suspect many of these companies are doing similar types of scraping, the largest and most aggressive on the conservative side being Cambridge Analytica … a sketchy (to say the least) data modelling company that has penetrated our market deeply." (Wong, 2019b). There was internal pressure from an employee who felt the case should be further investigated. However, the seriousness of the case appeared to be minimal when after a week no one responded. Eventually a response came that stated: "It's very likely these companies are not in violation of any of our terms," and "If we had more resources, we could discuss a call with the companies to get a better understanding, but we should only explore that path if we do see red flags.". Further discussion on the case lagged after that message, and employees concentrated on other matters (Wong, 2019b). The internal pressure was not strong enough to address the issue to senior executives. As a result, the pressure did not lead to any changes. It was only after the reveal in 2015, and later in 2018 that Facebook took serious measures.

Furthermore, after the public shock in 2018 more internal pressure came from employees. Alex Stamos, security official at Facebook, wrote a letter to his colleagues saying that the company had to take responsibility for the privacy scandal. In addition, he addressed the importance of internal changes to regain the trust of their users (Mac & Warzel, 2018).

Fake news and misinformation

After this event, internal pressures also came from within the company. From an interview with employees of Facebook, it appears that employees had already raised the issue of fake news with senior managers and top executives before the public shock. However, no response came from that other than 'they are working on it' (Frenkel, 2016).

After the public shock, more internal pressure came from employees. Employees disagreed with Mark Zuckerberg's statement that it was 'crazy' to think that fake news on Facebook influenced the election. In an interview with BuzzFeed, one employee said "It's not a crazy idea. What's crazy is for him to come out and dismiss it like that, when he knows, and those of us at the company know, that fake news ran wild on our platform during the entire campaign season." In response, employees created an unofficial task force to look at Facebooks' responsibility in countering fake news around the election and how their company acted. In response, Facebook updated its statement in which they no longer use the word 'crazy' (Frenkel, 2016).

Later, but in relation to elections and fake news, employees once again expressed their dissatisfaction about the company's decisions on this topic. Facebook decided to change a policy on political advertisement. The policy reads that posts and advertisements by politicians will not be removed, even if the content contains false information. Facebook made this decision because they do not want to censor political speech. However, Facebook employees do not agree with this decision. They argue that Facebook should not promote the dissemination of fake news. In addition, they find that every user, including politicians, should abide by the same rules of Facebook. That means that content involving fake news will be removed from the platform, regardless of an individual's political status (Isaac, 2020). However, the internal pressure did not result in change. In a response, Facebook said that the platform will not monitor political ads on fake information (Isaac & Kang, 2020).

Both cases show that internal pressures were present. However, from the analysis it appears that internal pressure by itself does not cause change. This is also evident in the example where Facebook employees pressured the company to change its new policy on political ads for the 2020 mid-term elections. To this, Facebook has said it is not going to change it. In this case, there has not been a public shock yet. This is also apparent in both cases regarding the 2016 elections. Concerns and pressures for change had been expressed by employees prior to the public shock. But these were at that time ignored. Only after the public shock did Facebook change its policy regarding data privacy, and fake news and misinformation. From this it seems that the public shock strengthens the internal pressures that can lead to change.

Summary

Analysis show that in both cases of the public shock of the US 2016 elections, there was pressure from internal and external factors prior to the public shock. Though, these were not taken seriously by Facebook. This resulted in a lack of extensive measures at that time. After the public shock, external pressure came from the media, governments, and political actors such as the European Union. Internal pressure came in both cases from employees who expressed their dissatisfaction with the company. From the analysis it seems that the public shock was important to strengthen the external and internal pressures in both Cambridge Analytica and the fake news and misinformation case to cause change.

In response to internal and external pressure following the public shock of the 2016 US election, Facebook has made changes to their privacy policies and fake news content moderation.

Moreover, technological improvements were made, as well as hiring additional employees in the safety and security domains.

In the case of Cambridge Analytica, Facebook released several statements saying they were going to tighten app permissions and privacy data controls on its platform. They have implemented stricter rules regarding data transfer to apps upon log-in, how apps use this data, and the removal of data following activity. Moreover, Facebook changed and improved their privacy settings by making it clear to users which apps have access to your data and simplifying these settings. These new measures are extensive, since Facebook has decided to apply these to its other tools such as Facebook Messenger, and to their other company, Instagram, as well (Schroepfer, 2018). Facebook has also changed certain policies because they are mandated by states and/or the European Union. To comply with this legislation, the changes were initially implemented regionally. However, these changes may eventually be extensive. In a Q&A, Zuckerberg said that he wants to apply the regulations set by the GDPR globally (Facebook, 2018d).

In response to external and internal pressures stemming from fake news, Facebook has implemented stricter rules for ads, entered into partnerships, and improved its machine learning to combat fake news and foreign interference. Since Facebook is used as a tool to reach voters not only during the U.S. elections, but also in many other countries, the new policies concerning ads and fake news during elections apply globally. Moreover, these policies, and the new technologies to counter fake news, also apply beyond election time (Fioretti, 2018). In addition, Facebook has entered into partnerships with press/news agencies in different countries to identify fake news. Facebook does this in several countries, and intend to expand this (Facebook, 2018c). Furthermore, the stricter and new regulations apply not only to Facebook, but also to partner networks Instagram and Messenger (Facebook, 2018e).

It is difficult to analyse in this case study whether external or internal factors led to change. This is because the external and internal pressures emerged almost simultaneously and alternated and reinforced each other. However, external pressure from the media is often followed by Facebook reacting about their actions. Nevertheless, in this case study, that does not necessarily mean that external pressures were more important. It does suggest that the public shock has strengthened both pressures to influence change.

Comparing the three cases

The analysis shows that in all three cases internal and/or external pressure had already occurred prior to the public shock. However, this was often ignored or not taken seriously by Facebook. Although these pressures were not influential at first, after the public shock, the internal and external pressure were important for change.

The analysis shows that, in all three cases, Facebook acted more adequately and extensively due to internal and external pressures following the public shock. The changes made were often profound since Facebook also extended its policies to other networks of their own, such as Facebook Messenger and Instagram. In addition, in all three cases, the new and updated content moderation policies are applied in a broader perspective than solely to the specific topic of the public shock and country. It has led to the broader adjustments globally and widely within Facebook

In Myanmar, for example, the new content moderation policy on white supremacy applies worldwide, and the stricter policies regarding hate speech apply to all sorts of hate speech. The regulations implemented after the Christchurch Call also apply globally. Furthermore, the changes to data privacy and fake news apply not only to US elections, but also to elections in other countries. In addition, the measures concerning fake news do not only relate to fake news during election time but apply to fake news in general. Facebook's broad adaptation is partly because the specific cases caused global comments and criticism from civil rights groups, governments, the media, and employees of Facebook.

However, the analysis also shows that Facebook does not always follow what internal and external factors requires of them. For example, the external pressure that was expressed to shut down Facebook Live was not followed up by Facebook. Nevertheless, Facebook did react to it by taking other stricter measures. This suggests that public shocks strengthened these pressures to become more influential.

Moreover, there is a difference whether external or internal pressures were more evident per case. In all cases external pressures have been evident. In all three cases, the media had played an important role. Moreover, the pressure from governments and political actors were most important in the cases. Facebook specifically acknowledges that these factors have influenced the stricter and new regulations implemented after the Christchurch attacks. Furthermore, Facebook responds to the external factors in the Myanmar case and the 2016 US election case by acknowledging their slow and ineffective response.

Internal pressures were also present, but the impact differed on how strong they were and whether they were direct or indirect. In Myanmar, the internal pressure came after the first regulations were announced and implemented. Moreover, pressure was not directly related to the case in Myanmar. Therefore, the internal pressures did not directly lead to change. Furthermore, internal pressure also came after the Christchurch attacks. The analysis shows that only advertisers exerted internal pressure on Facebook. However, the effect of this internal factor alone is not considered to be very significant, as Facebook experienced much more external pressure from civil rights groups and the government. In these two cases, it seems that, compared to external pressure, internal pressure was not the main factor that led to change. In the case of the 2016 US elections, however, internal pressures may have had a greater impact. In this case, there was a lot of pressure from employees regarding Facebook's direction and decisions. In response, Facebook changed its regulations on data privacy and on the content moderation of fake news and misinformation. However, external factors were probably also influential here.

In general, it is more difficult to know for sure whether internal pressures lead to change. This is because in the case of external pressures, Facebook often gives a reaction to the media or on their website. Furthermore, it is hard to be certain if either external or internal pressures led to change. As mentioned before, the two are not distinct. In fact, internal and external factors can reinforce each other, leading both to content moderation change. However, since I distinguish the internal and external pressure in this thesis, the analysis shows that external pressures are the most influential to cause change.

Conclusion

To conclude, this research has used a case-oriented, qualitative, process-tracing methodology centred on Facebook's content moderation. This research adopted the concept of public shocks by Ananny and Gillespie (2017). Using their concept, I studied if internal factors or external factors led to content moderation change at social media companies after a public shock.

To study this, my research question reads: *How do moments of public shock shape the content moderation policies of social media?*

To answer the research question, I conducted a case study of three public shocks that happened at Facebook: Myanmar, the Christchurch attacks, and the 2016 US election. On the basis of four sub questions, I analysed whether internal or external factors were apparent in each case and assessed the depth of change in content moderation policies.

The four sub questions helped me to make a structured analysis per case. For each case, I examined whether internal and/or external pressure occurred after the public shock, and whether this led to a change in Facebook's content moderation policies. I also examined whether there had been pressure prior to a public shock and how Facebook had reacted to it. In addition, I looked at whether the changes implemented were extensive. In this way, the depth of the change that occurs after these shocks is shown.

From the analysis of the three cases follows that external pressures from the media, governments, political actors, and civil rights groups were most important. Since Facebook is one of the largest platforms, I carefully make the generalization that this could apply to social media companies in general. Therefore, the answer to the research question is: content moderation policies of social media are mainly shaped by external pressures after a public shock.

This research shows that external factors have a great influence on the content moderation change of social media companies. This indicates that social media companies should be approached from this perspective from now on if one wants to change them. However, further research should be done to determine the most efficient external strategy to this approach.

Moreover, from the analysis follows that it is more difficult to obtain evidence from Facebook whether internal pressures have led to change than in the case of external pressure. Therefore, research that focus specifically on the internal factors would provide more insight into its impact. Furthermore, further research can focus on whether the same factors have led to change in other large social media companies such as Twitter or YouTube. Lastly, further research can be done on whether other (large) social media companies have been influenced by Facebook's changes to content moderation.

Reference list

Allcott, H., & Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives*, *31*(2), 211–236. https://doi.org/10.1257/jep.31.2.211

Amer, K., & Noujaim, J. (Director). (2019). *The Great Hack* [Film]. Netflix.

Ananny, M., and T. Gillespie. (2017). "Public Platforms: Beyond the Cycle of Shocks and Exceptions." *Oxford Internet Institute*. September 8. http://blogs.oii.ox.ac.uk/ippconference/sites/ipp/files/documents/anannyGillespie-publicPlatforms-oiisubmittedSept8.pdf

Archibong, I. (2019, September 20). *An Update on Our App Developer Investigation*. About Facebook. https://about.fb.com/news/2019/09/an-update-on-our-app-developer-investigation/

Beach, D., & Pedersen, R. B. (2016). *Causal Case Study Methods: Foundations and Guidelines for Comparing, Matching, and Tracing*. University of Michigan Press.

Beales, H., Brito, J., Kennerly Davis, J., DeMuth, C., Devine, D., Dudley, S., Mannix, B., & McGinnis, J. O. (2017). Government Regulation: The Good, The Bad, & The Ugly. *Regulatory Process Working Group*, 3–17. https://regproject.org/paper/government-regulation-the-good-the-bad-the-ugly/

Bickert, M., & Fishman, B. (2017, June 15). *Hard Questions: How We Counter Terrorism*. About Facebook. https://about.fb.com/news/2017/06/how-we-counter-terrorism/

Bond, D., & Robinson, D. (2017, January 30). European Commission fires warning at Facebook over fake news. *Financial Times*. https://www.ft.com/content/85683e08-e4a9-11e6-9645-c9357a75844a

Cadwalladr, C., & E. Graham-Harrison, E. (2018, March 17). *Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach*. The Guardian. https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election

Christchurch Call. (2021). *Christchurch Call Community Consultation. Final Report*. https://www.christchurchcall.com/christchurch-call-community-consultation-report.pdf

Christchurch Call. (n.d.). *Christchurch Call | to eliminate terrorist and violent extremist content online*. Retrieved July 6, 2021, from https://www.christchurchcall.com/

Citron, D. K., & Franks, M. A. (2020). The Internet As a Speech Machine and Other Myths Confounding Section 230 Speech Reform. *University of Chicago Legal Forum*, *2020*(3), 45–75. https://doi.org/10.2139/ssrn.3532691

Coldewey, D. (2020, October 19). *Who regulates social media? Good question!* Tech Crunch. https://tinyurl.com/5x8jzhey

Community Standards. (2021). Facebook. https://www.facebook.com/communitystandards/introduction

Confessore, N. (2018, 15 November). *Cambridge Analytica and Facebook: The Scandal and the Fallout So Far*. The New York Times. https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html

Cox, J. (2018, May 29). *These Are Facebook's Policies for Moderating White Supremacy and Hate*. Vice. https://www.vice.com/en/article/mbk7ky/leaked-facebook-neo-nazi-policies-white-supremacy-nationalism-separatism

Cox, J., & Koebler, J. (2018, September 20). *Facebook Is Reviewing its Policy on White Nationalism After Motherboard Investigation, Civil Rights Backlash*. Vice. https://www.vice.com/en/article/yw4pbj/facebook-white-supremacy-white-nationalism-hate-speech-policy

Cox, J., & Koebler, J. (2019, March 27). *Facebook Bans White Nationalism and White Separatism*. Vice. https://www.vice.com/en/article/nexpbx/facebook-bans-white-nationalism-and-white-separatism

Davies, H. (2015, December 11). *Ted Cruz using firm that harvested data on millions of unwitting Facebook users*. The Guardian. https://www.theguardian.com/us-news/2015/dec/11/senator-ted-cruz-president-campaign-facebook-user-data

De Streel, A., Defreyne, E., Jacquemin, H., Ledger, M., & Michel, A. (2020, June). *Online Platforms' Moderation of Illegal Content Online*. Study for the committee on Internal Market and Consumer Protection, Policy Department for Economic, Scientific and Quality of Life Policies, European Parliament. https://www.europarl.europa.eu/RegData/etudes/STUD/2020/652718/IPOL_STU(2020)652718_EN.pdf

Edmunds, S. (2019, March 18). *Lotto, Westpac, TSB pull online adverts in the wake of the Christchurch shootings*. Stuff. https://www.stuff.co.nz/national/christchurch-shooting/111379772/new-zealand-advertisers-reconsider-social-media-in-wake-of-christchurch-attacks

Egan, E., & Beringer, A. (2018, March 28). *It's Time to Make Our Privacy Tools Easier to Find*. About Facebook. https://about.fb.com/news/2018/03/privacy-shortcuts/

Facebook. (2018a, January 29). *Facebook's Commitment to Data Protection and Privacy in Compliance with the GDPR*. Facebook for Business. https://www.facebook.com/business/news/facebooks-commitment-to-data-protection-and-privacy-in-compliance-with-the-gdpr

Facebook. (2018b, March 21). *Cracking Down on Platform Abuse*. About Facebook. https://about.fb.com/news/2018/03/cracking-down-on-platform-abuse/

Facebook. (2018c, March 29). *Hard Questions: What is Facebook Doing to Protect Election Security?* About Facebook. https://about.fb.com/news/2018/03/hard-questions-election-security/

Facebook. (2018d, April 4). *Hard Questions: Q&A With Mark Zuckerberg on Protecting People's Information*. About Facebook. https://about.fb.com/news/2018/04/hard-questions-protecting-peoples-information/

Facebook. (2018e, June 28). *Q&A on Ads and Pages Transparency*. About Facebook. https://about.fb.com/news/2018/06/qa-on-ads-and-pages-transparency/

Facebook. (2018f, August 28). *Removing Myanmar Military Officials From Facebook*. About Facebook. https://about.fb.com/news/2018/08/removing-myanmar-officials/

Facebook. (2019a, March 27). *Standing Against Hate*. About Facebook. https://about.fb.com/news/2019/03/standing-against-hate/

Facebook. (2019b, June). *Facebook's Civil Rights Audit – Progress Report*. https://about.fb.com/wp-content/uploads/2019/06/civilrightaudit_final.pdf

Facebook. (2019c, September 17). *Combating Hate and Extremism*. About Facebook. https://about.fb.com/news/2019/09/combating-hate-and-extremism/

Facebook. (2019d, November 7). *Facebook Joins Other Tech Companies to Support the Christchurch Call to Action*. About Facebook. https://about.fb.com/news/2019/05/christchurch-call-to-action/

Facebook. (2020, May 15). *An Update on Combating Hate and Dangerous Organizations*. About Facebook. https://about.fb.com/news/2020/05/combating-hate-and-dangerous-organizations/

Fink, C. (2018). Dangerous Speech, Anti-Muslim Violence, and Facebook in Myanmar. *Journal of International Affairs, Special Issue*, *71*(1,5), 43–52. https://jia.sipa.columbia.edu/dangerous-speech-anti-muslim-violence-and-facebook-myanmar

Fioretti, J. (2018, April 26). EU piles pressure on social media over fake news. *Reuters*. https://www.reuters.com/article/us-eu-internet-fakenews-idUSKBN1HX15D

Fisher, M. (2018, December 27). Inside Facebook’s Secret Rulebook for Global Political Speech. *The New York Times*. https://www.nytimes.com/2018/12/27/world/facebook-moderators.html

Flew, T., Martin, F., & Suzor, N. (2019). Internet regulation as media policy: Rethinking the question of digital communication platform governance. *Journal of Digital Media & Policy*, *10*(1), 33–50. https://doi.org/10.1386/jdmp.10.1.33_1

Frankel, R. (2021, April 15). *An Update on the Situation in Myanmar*. About Facebook. https://about.fb.com/news/2021/02/an-update-on-myanmar/

Frenkel, S. (2016, November 14). *Renegade Facebook Employees Form Task Force To Battle Fake News*. BuzzFeed News. https://www.buzzfeednews.com/article/sheerafrenkel/renegade-facebook-employees-form-task-force-to-battle-fake-n

Gallo, J. A., & Cho, C. Y. (2021, January). *Social Media: Misinformation and Content Moderation Issues for Congress*. Congressional Research Service. https://www.everycrsreport.com/reports/R46662.html

Gill, K. (2021). Regulation Platforms’ Invisible Hand: Content Moderation Policies and Processes. *Journal of Business & Intellectual Property Law, 21*(2), 171–212. https://heinonline.org/HOL/Page?collection=journals&handle=hein.journals/wakfinp21&id=186&men_tab=srchresults

Gillespie, T. (2018a). *Custodians of the Internet*. Yale University Press.

Gillespie, T. (2018b). Platforms Are Not Intermediaries. *Georgetown Law Technology Review, 2*(2), 198–216. https://resources.platform.coop/resources/platforms-are-not-intermediaries/

Grewal, P. (2018, March 16). *Suspending Cambridge Analytica and SCL Group From Facebook*. About Facebook. https://about.fb.com/news/2018/03/suspending-cambridge-analytica/

Gunia, A. (2019, 15 May). Facebook Tightens Live-Stream Rules in Response to the Christchurch Massacre. *Time*. https://time.com/5589478/facebook-livestream-rules-new-zealand-christchurch-attack/

Hoekstra, A. (2020, 23 June). Facebook nog altijd spil in haatpropaganda tegen Rohingya. *Trouw*. https://www.trouw.nl/buitenland/facebook-nog-altijd-spil-in-haatpropaganda-tegen-rohingya~b57385c6/?referrer=https%3A%2F%2Fwww.google.com%2F

Human Rights Watch. (2019). *World Report 2019: Myanmar. Events of 2018*. https://www.hrw.org/world-report/2019/country-chapters/myanmar-burma#

Ingram, D. (2017, May 3). *Facebook tries to fix violent video problem with 3,000 new workers*. Reuters. https://www.reuters.com/article/us-facebook-crime-idUSKBN17Z1N4

Isaac, M. (2020, July 10). *Dissent Erupts at Facebook Over Hands-Off Stance on Political Ads*. The New York Times. https://www.nytimes.com/2019/10/28/technology/facebook-mark-zuckerberg-political-ads.html

Isaac, M., & Kang, C. (2020, September 4). *Facebook Says It Won't Back Down From Allowing Lies in Political Ads*. The New York Times. https://www.nytimes.com/2020/01/09/technology/facebook-political-ads-lies.html

Isaac, M., & Wakabayashi, D. (2017, October 31). *Russian Influence Reached 126 Million Through Facebook Alone*. The New York Times. https://www.nytimes.com/2017/10/30/technology/facebook-google-russia.html

Klonick, K. (2018). The New Governors: The People, Rules, and Processes Governing Online Speech. *Harvard Law Review, 131*(6), 1599-1670. https://www.researchgate.net/publication/324645451_The_new_governors_The_people_rules_and_processes_governing_online_speech

Klonick, K. (2019, April 25). *Inside the Team at Facebook That Dealt with the Christchurch Shooting*. The New Yorker. https://www.newyorker.com/news/news-desk/inside-the-team-at-facebook-that-dealt-with-the-christchurch-shooting

Klonick, K. (2020). The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression. *The Yale Law Journal*, 2418–2499. https://heinonline.org/HOL/Page?handle=hein.journals/ylr129&id=2476&collection=journals&index=

Koenis, C. (2020, October 13). *Waarom de invloed van Facebook op de Amerikaanse verkiezingen groot blijft*. RTL Nieuws. https://www.rtlnieuws.nl/nieuws/buitenland/artikel/5189856/facebook-inmenging-verkiezingen-vs-trump-biden-nepaccounts

Koeze, E., & Popper, N. (2020, 8 April). *The Virus Changed the Way We Internet*. The New York Times. https://www.nytimes.com/interactive/2020/04/07/technology/coronavirus-internet-use.html

Lessig, L. (2009). *Code 2.0*. Van Haren Publishing.

Levin, S. (2017, September 28). Mark Zuckerberg: I regret ridiculing fears over Facebook's effect on election. *The Guardian*. https://www.theguardian.com/technology/2017/sep/27/mark-zuckerberg-facebook-2016-electio2016-election-fake-news

Lewis, J. A. (2017, November 1). *European Union to Social Media: Regulate or Be Regulated*. Center for Strategic and International Studies. https://www.csis.org/analysis/european-union-social-media-regulate-or-be-regulated

Lomas, N. (2018, April 10). *How Facebook has reacted since the data misuse scandal broke*. Tech Crunch. https://techcrunch.com/2018/04/10/how-facebook-has-reacted-since-the-data-misuse-scandal-broke/

Mac, R., & Warzel, C. (2018, July 24). *Departing Facebook Security Officer's Memo: "We Need To Be Willing To Pick Sides."* BuzzFeed News. https://www.buzzfeednews.com/article/ryanmac/facebook-alex-stamos-memo-cambridge-analytica-pick-sides

Macklin, G. (2019). The Christchurch Attacks: Livestream Terror in the Viral Video Age. *CTC Sentinel*, *12*(6), 18–29. https://ctc.usma.edu/christchurch-attacks-livestream-terror-viral-video-age/

Maroni, M. (2020, 29 June). *Some reflections on the announced Facebook Oversight Board*. Centre for Media Pluralism and Freedom. https://cmpf.eui.eu/some-reflections-on-the-announced-facebook-oversight-board/

Mazúr, J., & Patakyová, M. T. (2019). Regulatory Approaches to Facebook and Other Social Media Platforms: Towards Platforms Design Accountability. *Masaryk University Journal of Law and Technology, 13*(2), 219–242. https://doi.org/10.5817/mujlt2019-2-4

Metz, C., & Satariano, A. (2019, May 14). *Facebook Restricts Live Streaming After New Zealand Shooting*. The New York Times. https://www.nytimes.com/2019/05/14/technology/facebook-live-violent-content.html

Miles, T. (2018, 12 March). U.N. investigators cite Facebook role in Myanmar crisis. *Reuters*. https://www.reuters.com/article/us-myanmar-rohingya-facebook/u-n-investigators-cite-facebook-role-in-myanmar-crisis-idUKKCN1GO2PN?edition-redirect=uk

Mozur, P. (2018, 15 October). A Genocide Incited on Facebook, With Posts From Myanmar's Military. *The New York Times*. https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html

Napoli, P. M. (2019). User Data as Public Resource: Implications for Social Media Regulation. *Policy & Internet, 11*(4), 439–459. https://doi.org/10.1002/poi3.216

NOS. (2018, 17 March). Gegevens 50 miljoen kiezers VS via Facebook buitgemaakt. *NOS*. https://nos.nl/artikel/2222991-gegevens-50-miljoen-kiezers-vs-via-facebook-buitgemaakt

NU.nl. (2018, 10 April). Zo werd Facebook-data mogelijk misbruikt door Trump-campagne. *NU*. https://www.nu.nl/internet/5182130/zo-werd-facebook-data-mogelijk-misbruikt-trump-campagne.html

Open hearing: Social Media Influence in the 2016 U.S. Elections: Hearing beforte the select Committee on Intelligence of the U.S. Senate, 115th Cong. (2017) https://www.govinfo.gov/content/pkg/CHRG-115shrg27398/pdf/CHRG-115shrg27398.pdf

Oversight Board. (2020). Oversight Board Charter. Facebook Oversight Board. https://oversightboard.com/

Paul, K., & Dang, S. (2020, June 26). *Verizon suspends advertising on Facebook, joins growing boycott*. Reuters. https://www.reuters.com/article/us-facebook-ads-boycott-verizon-idUSKBN23W3HK

Perrigo, B. (2021, June 24). *Facebook Tried to Ban Myanmar's Military. But Its Own Algorithm Kept Promoting Pages Supporting Them, Report Says*. Time. https://time.com/6075539/facebook-myanmar-military/

Potkin, F. (2021, February 4). *Facebook faces a reckoning in Myanmar after blocked by military*. Reuters. https://www.reuters.com/article/us-myanmar-politics-facebook-focus-idUSKBN2A42RY

Rajagopalan, M., Vo, L. T., & Soe, A. N. (2018, 15 oktober). *How Facebook Failed The Rohingya In Myanmar*. BuzzFeed News. https://www.buzzfeednews.com/article/meghara/facebook-myanmar-rohingya-genocide

Reuters Staff. (2020, July 3). *Factbox: More companies join Facebook ad boycott bandwagon*. Reuters. https://www.reuters.com/article/us-facebook-ads-boycott-factbox-idUSKBN2433CL

Roose, K., & Mozur, P. (2018, April 10). *Zuckerberg Was Called Out Over Myanmar Violence. Here's His Apology.* The New York Times. https://www.nytimes.com/2018/04/09/business/facebook-myanmar-zuckerberg.html

Rosen, G. (2019, November 7). *Protecting Facebook Live From Abuse and Investing in Manipulated Media Research*. About Facebook. https://about.fb.com/news/2019/05/protecting-live-from-abuse/

Sandberg, S. (2019, March 30). *Facebook Chief Operating Officer Sheryl Sandberg's letter to New Zealand*. The New Zealand Herald. https://www.nzherald.co.nz/business/facebook-chief-operating-officer-sheryl-sandbergs-letter-to-new-zealand/UAPCQMTI645ICB734DKQW25FQQ/

Schroepfer, M. (2018, April 4). *An Update on Our Plans to Restrict Data Access on Facebook*. About Facebook. https://about.fb.com/news/2018/04/restricting-data-access/

Seetharaman, D. (2017, March 6). *Facebook, Rushing Into Live Video, Wasn't Ready for Its Dark Side*. The Wall Street Journal. https://www.wsj.com/articles/in-rush-to-live-video-facebook-moved-fast-and-broke-things-1488821247

Shead, B. S. (2019, 18 December). *Facebook owns the four most downloaded apps of the decade*. BBC News. https://www.bbc.com/news/technology-50838013

Simon, M. K., & Goes, J. (2010). *Dissertation & Scholarly Research: Recipes for Success* (2nd ed.). CreateSpace Independent Publishing Platform.

Slodkowski, A. (2018, August 27). Facebook bans Myanmar army chief, others in unprecedented move. *Reuters*. https://www.reuters.com/article/us-myanmar-facebook/facebook-removes-pages-of-top-myanmar-military-official-others-idUSKCN1LC0R7

Smith, M. (2020, 18 August). *Facebook Wanted to Be a Force for Good in Myanmar. Now It Is Rejecting a Request to Help With a Genocide Investigation*. Time. https://time.com/5880118/myanmar-rohingya-genocide-facebook-gambia/

Sonderby, C. (2019, November 7). *Update on New Zealand*. About Facebook. https://about.fb.com/news/2019/03/update-on-new-zealand/

Stecklow, S. (2018, August). *Inside Facebook's Myanmar operation Hatebook. Why Facebook is losing the war on hate speech in Myanmar.* Reuters. https://www.reuters.com/investigates/special-report/myanmar-facebook-hate/

Su, S. (2018, August 15). *Update on Myanmar*. About Facebook. https://about.fb.com/news/2018/08/update-on-myanmar/

Thomas, G. (2011). A Typology for the Case Study in Social Science Following a Review of Definition, Discourse, and Structure. *Qualitative Inquiry*, *17*(6), 511–521. https://doi.org/10.1177/1077800411409884

Thompson, N., & Vogelstein, F. (2018, March 20). *Facebook Struggles to Respond to the Cambridge Analytica Scandal*. Wired. https://www.wired.com/story/facebook-cambridge-analytica-response/

Timberg, C., & Dwoskin, E. (2020, September 8). *Another Facebook worker quits in disgust, saying the company 'is on the wrong side of history.'* The Washington Post.

https://www.washingtonpost.com/technology/2020/09/08/facebook-employee-quit-racism/

United Nations Human Rights Council. (2018, September). *Report of the independent international fact-finding mission on Myanmar*. https://www.ohchr.org/Documents/HRBodies/HRCouncil/FFM-Myanmar/A_HRC_39_64.pdf

Van Bemmel, N. (2020, May 1). Facebooks strijd tegen nepnieuws: 'Je merkt dat ze worstelen met hun rol.' *De Volkskrant*. https://www.volkskrant.nl/wetenschap/facebooks-strijd-tegen-nepnieuws-je-merkt-dat-ze-worstelen-met-hun-rol~b4641bf6/

Van Dijck, J. (2013). *The Culture of Connectivity: A Critical History of Social Media.* Oxford University Press.

Warofka, A. P. P. M. (2018, 5 November). *An Independent Assessment of the Human Rights Impact of Facebook in Myanmar*. About Facebook. https://about.fb.com/news/2018/11/myanmar-hria/

Watson, C. (2018, April 11). *The key moments from Mark Zuckerberg's testimony to Congress*. The Guardian. https://www.theguardian.com/technology/2018/apr/11/mark-zuckerbergs-testimony-to-congress-the-key-moments

Wichter, Z. (2018, April 12). *2 Days, 10 Hours, 600 Questions: What Happened When Mark Zuckerberg Went to Washington*. The New York Times. https://www.nytimes.com/2018/04/12/technology/mark-zuckerberg-testimony.html

Wolford, B. (n.d.). *What is GDPR, the EU's new data protection law?* GDPR.EU. Retrieved July 31, 2021, from https://gdpr.eu/what-is-gdpr/

Wong, J. C. (2019a, March 22). *Facebook acknowledges concerns over Cambridge Analytica emerged earlier than reported*. The Guardian. https://www.theguardian.com/uk-news/2019/mar/21/facebook-knew-of-cambridge-analytica-data-misuse-earlier-than-reported-court-filing

Wong, J. C. (2019b, 24 August). Document reveals how Facebook downplayed early Cambridge Analytica concerns. *The Guardian*. https://www.theguardian.com/technology/2019/aug/23/cambridge-analytica-facebook-response-internal-document