Internet Governance and Online Identity:

Searching for Correlations between User Anonymity and User Behaviour

by

Eric Robert Silver

Submitted to the Faculty of Arts

University of Alberta

in partial fulfillment of the requirements for the degree of

Master of Arts in Communication and Technology

August 13, 2021

Acknowledgements

I would like to thank all the professors, instructors, and support staff from the Master of Arts in Communications and Technology (MACT) program at the University of Alberta, especially my Capstone supervisor, Dr. Rob McMahon. Dr. McMahon's support and guidance kept this project on track and moving forward and was greatly appreciated during this (at times) challenging process.

I would also like to acknowledge the incredible group of people within our 2019 cohort who have been a continuous source of support throughout this program. Specifically, I'd like to recognize a few members who were a valued sounding board, working group, and social circle: Sheena Fitzpatrick, Michelle Stevenson, and Candace Westeroth.

Finally, my deepest thanks go to my wife for motivating me to sit down, read, write, and soldier through this project. This wouldn't have been possible without you; I love you.

Contents

Tables and Figures

Abstract

Anonymity can grant people privacy, safety, and enable rebellion in an oppressive regime. But it can also foster aggression, hate, and undue political upheaval. Drawing on theoretical frameworks from Science and Technology Studies, this research aims to learn how user anonymity granted by specific platform structures and policies influences how users behave either prosocially or antisocially online. To reach this goal, I performed a qualitative content analysis on data sampled from three major platforms with varying levels of user anonymity (Facebook, Reddit, and 4chan), studied each platform's basic structure, content, and privacy policies, and researched government policies and procedures about online anonymity in Canada and abroad. My analysis focused on collecting a sample of one hundred comments from one root post on each platform, and then conducting an inductive content analysis on those data sets. One of this study's primary findings supports existing literature that states that levels of anonymity directly influence user behaviour online, in both prosocial and antisocial ways. Within my research, users of 4chan, the platform design with the highest level of anonymity, produced the highest number of aggression codes and the most severe aggressive behaviour, yet also the least user-directed aggression, the highest number of non-aggression posts (tied), and the most varied non-aggression codes. This finding leads to further speculation that online anonymity can neither be considered helpful or harmful;it is a complex social construct that brings out both the best and the worst of human behaviour.

Chapter 1: Introduction

Within this exploratory Capstone research project, I study relationships between online anonymity, internet governance (digital platform structure and government policy) and user behaviour. Through a qualitative content analysis, I hope to answer the following research questions: "How do anonymity, policy, and structure affect social media users' behaviour online, and how does user behaviour in turn shape platform structure?"

This research project draws from literature within the field of Science and Technology Studies, and specifically the theory of Social Shaping of Technology (SST), which looks at how users influence technology, as well as how technology influences users. SST recognizes that "people, technologies, and institutions all have power to influence the development and subsequent use of technology" and that "machines can and do accelerate certain trends, cultural weaknesses, and fortify certain social structures while eroding others" (Baym, 2105, pp 51-52). This theory sees the relationships between technology, users, and society as cyclical, with each actor influencing and being influenced by the other. My project will explore how the structural design of technical platforms influences how users behave online, as well as how human actors such as these users and government policy makers influence the design of platforms.

I believe that significant social benefits can be gained from research on these issues. Online behaviour can directly impact offline behaviour, often with severe outcomes. For example, a 15-year-old girl's suicide in British Columbia, Canada in 2012 garnered international media attention (BBC, 2017; CTV, 2012; Dean, 2012; Miljure and Mangione, 2021). Amanda Todd was harassed relentlessly by a person she met online who was posing as a friendly, flattering man. When she would not cooperate with his perverse demands, he shared a private photo of her with classmates, friends, and family members through online messages. This

harassment continued as she moved schools, and eventually caused her to take her own life, as revealed in a video message she left behind (TheSomebodytoknow, 2012). The person harassing Amanda Todd was hiding behind an anonymous account, which created in incredible challenge for authorities to find and prosecute the individual to meet the community's demands for justice. Various tips from online sources and even the hacktivist group Anonymous pointed the finger at numerous men believed to be this anonymous figure (Beck, 2013; CTV BC, 2012). However, these accusations were proven incorrect, and it was not until 2014 that the true perpetrator, a Dutch man named Aydin Coban, was apprehended – and not until 2021 was he extradited to Canada for trial. Coban was previously sentenced by a Dutch court to 11 years in prison after being convicted of similar crimes against 34 other young people (Miljure and Mangione, 2021). This example shows how one user's online anonymity can have incredible impacts on the offline lives of others. The impacts extend not just to Amanda Todd, the victim of his cyber aggression, and her family, but also to the men who were falsely accused, as well as all other victims who may have been protected had he been apprehended sooner.

Despite this and other high profile examples of the antisocial outcomes of online anonymity, I learned in this research project that the relationship between anonymity and user behaviour is not as straightforward as a simple cause-and-effect. In fact, anonymity can produce both prosocial and antisocial effects. While not as prominently researched, many sources in my literature review praise anonymity as the cause of prosocial behaviour. Seeing both sides of this conversation led me to question how anonymity is made available to users online. Different government policies and platform design choices give varying degrees of anonymity to users, which in turn influences how they behave online. Researching how the internet is governed, both by governments and platforms, may shed new light on how platform structures and anonymity

influence user behaviour online. This kind of research can help highlight this conversation in the public sphere, ideally leading to a safer user experience, both online and offline.

I begin my capstone with a systematic literature review of anonymity, user behaviour, digital platform structure, government online policy, and theoretical frameworks within the realm of Science and Technology Studies. To select an appropriate theoretical framework for my project, I looked at four prominent theories: Technological Determinism, Social Construction of Technology, Social Shaping of Technology, and Domestication of Technology. Whether it is being studied offline or online, many authors have looked at how anonymity influences society. In online contexts, the findings within the research literature are mixed but skew heavily towards agreeing that anonymity increases antisocial online posting behaviour more than prosocial. Because of the complexity and evolving nature of the digital sphere, there is no standard terminology to describe harms and levels of anonymity in online contexts. Some authors working in this space refer to major theories, such as Suler's (2004) online disinhibition effect, deindividuation, or the Social Identity Model of Deindividuation Effects (SIDE) (Lea et al., 2001). Others, who focus only on the prosocial effects of anonymity, may describe them as benign, helpful, cooperative, or positive. Antisocial effects have been called toxic, harmful, aggressive, flaming, and more. Within this research project, the terms "aggression" and "non-aggression" will be used when referring to online anonymity. The term "harmful" will be used as an umbrella to describe users' behaviour, rather than writing too granularly about each individual behaviour. However, each behaviour will be assigned an individual code such as "threats of violence," "racism," or "benign comment" during data analysis.

The review of the literature also demonstrated there are multiple ways to describe online anonymity, based on how identifiable a user is. This level of anonymity is often an inherent

technical aspect of a platform rather than a user choice. Pfitzmann and Köhntopp (2001) identified three terms for different levels of anonymity users can have online (anonymity, unobservability, and pseudonymity). Within the definition of pseudonymity, they differentiate between varying levels of links for these account types, such as public, linkable, and unlinkable (Pfitzmann & Köhntopp, 2001). These terms may be used within this research project, as well as terminology from other sources, such as nonymity (being known; the antonym of anonymity). However, for the purposes of this project, I define Facebook as the least anonymous, 4chan as the most anonymous, and Reddit as pseudonymous (mid-range anonymity) platform.

Next, to learn more about how these three online platforms govern issues related to anonymity and behaviour, I reviewed the content policies, privacy policies, and overall structures of Facebook, Reddit, and 4chan. I present a summary of these issues in a subsection of the literature review. I also looked at current Canadian government policies, as well as studies from other nations, to better understand how these issues are administered by governments.

Following the literature review section, I discuss my methodology and research design. Within this section, I provide an overview of my project design, including sampling, data collection, data storage, analysis, coding, and theoretical framework. To guide my data analysis process, I followed Saldaña's (2013) 32 methods of qualitative coding and used an open-source software, Taguette, to organize and code raw platform data.

The final section of this research project is a thorough discussion and analysis of my findings. This section provides examples of raw data and the codes used to categorize them into analyzable chunks. I use charts, tables, and images to illustrate my findings in an easily digestible way, and help show readers what I found, rather than just telling them. My discussion of these findings reflects on the literature about anonymity and user behaviour. After completing

the research and data analysis, it was clear that while anonymity may not have a large impact on the likelihood to post aggressive posts online, it does impact the type and severity of aggression that users display. Each platform's structure and policies also influenced the ways users behaved. Without giving too much away in the introduction, I will say that my findings support much of the existing literature that states that anonymity directly influences user behaviour online, both prosocially and antisocially.

Chapter 2: Literature Review

Introduction

I present this review of literature, policy, and theories applicable to my project in a way that provides steps for readers to replicate this research strategy and compile a systematic literature review. I grouped the literature thematically into three main sections: theoretical perspectives, the social effects of anonymity, and the current state of government and platform policies about internet governance.

This literature begins with a look at four theoretical frameworks that will inform this research project's design: Technological Determinism, Social Construction of Technology, Social Shaping of Technology, and Domestication of Technology. Following this section, I review overarching effects of anonymity, including sources that focus on online and offline anonymity, as well as toxic (antisocial) and benign (prosocial) effects. A thorough, systematic review of the research literature about anonymous user behaviour on digital platforms found that many researchers have studied the relationships between anonymity and social effects, and specifically harmful, or toxic, effects. The literature I reviewed in this area extends as far back as the 1960s, and as recently as 2020. There are also examples of studies that look at the prosocial, positive, or benign effects of anonymity, particularly in the digital sphere. Another area I researched was the current landscape around internet governance in terms of the policies of national governments and of digital platforms themselves. This section of the review aims to understand how the Canadian government handles online privacy and toxic behaviour. It also examines policies implemented by other nations, in order to identify connections between anonymity, user behaviour, and environment. This policy section also reviews the content policies, privacy policies, and user conduct policies of the three platforms studied – Facebook,

Reddit, and 4chan – in order to understand the structural and technical components that provide users different level of anonymity.

This literature review identifies a gap in existing research in the areas of internet governance, anonymity, and user behaviour is identified. While many studies aim to prove that degrees of anonymity influence behaviour in some way, there has not been any substantial work done comparing the different levels of anonymity present in multiple platforms' structures. Specifically, much of the existing literature studies how anonymity can cause either prosocial or antisocial behaviour; this body of work is captured within the second section of this literature review, The Significant Social Impacts of Anonymity. Also, many sources use a single platform to explore user behaviour, rather than comparing differences across multiple platforms (Ascher & Noble, 2019; Bernstein, 2011; Kurek, 2019; Mikal et al., 2016; Pourghomi et al., 2020; Siegfried-Spellar & Lankford, 2018; Young et al., 2018). Based on this gap, the research conducted within this Capstone project will compare user behaviour across three platforms with different anonymity structures to try and learn whether there are any relationships between levels of user anonymity and online behaviour. These findings will be synthesized in the project conclusion, along with a section that provides possible recommendations to government and platform decision-makers.

Literature Review Methodology

The literature review methodology employed multiple techniques to search the literature, including Boolean logic, bibliographic search, author search, colleague recommendations, and a catalogue search of the *Canadian Journal of Communication.* Details of these tactics are outlined below to facilitate the replication of the search process and therefore increase the validity of results.

The first tool used was the University of Alberta's online library, specifically the EBSCO database, limited to peer-reviewed articles. The Boolean terms used in this phase are listed in Table 1 below. There was an intentional omission of terms related to politics, terrorism, and disinformation/misinformation. These topics are very common in discussions of online harms, but this research project is not focused on these specific concerns.

Boolean Search Terms				
(online or virtual or digital)	AND	(anonym* or pseudonym* or unlinkab* or unobserv* or unperceiv*)	AND	("flaming behavio*r" or "harmful behavio*r" or cyberbully* or "hate speech" or "online harm*" or "online aggression")
(online or virtual or digital)	AND	(anonym* or pseudonym* or unlinkab* or unobserv* or unperceiv*)	AND	(disinhibit* or deindividuat* or de-individuat*)
("social media" or "digital platform")	AND	anon*	AND	policy
Canad*	AND	("online policy" or "digital policy" or "communication* policy")		

Table 1: Boolean Search Terms

In addition to this strategy, recommendations of material received from classmates and instructors led to many of the same articles identified through the Boolean search, as well as a number of previously unidentified ones as well. Reviewing the references in identified literature and performing a bibliographic search led to even more relevant sources, and performing an author search on some key articles proved fruitful as well. Finally, a database search of the *Canadian Journal of Communication* was performed, specifically looking for the term "policy".

This strategy proved to be the least successful of all searches, producing only one relevant article.

Theoretical Review and Framework

  The final section of this literature review looks into four related theories that could potentially form the framework for this research project. Drawn from Science and Technology Studies, these four theories provide explanations of the relationships between technology and users/society. In her 2015 book, *Personal Connections in the Digital Age*, Nancy Baym highlights these four theories in a chapter titled "Making new media make sense." This section of the literature review summarizes her work and borrows insight from other authors to provide an overview of each of these theories that may guide this research project. There are four theoretical assumptions that can be made about the way technology influences society. One is Technological Determinism, which argues that technologies are "active forces of change that humans have little power to resist" (Baym, 2015, p. 26). The second is the Social Construction of Technology, which claims that "people are the primary sources of change in both technology and society" (Baym, 2015, p. 26). The third theory, which my research project is most closely informed by, is the Social Shaping of Technology, which "sees technology and society as continually influencing one another" (Baym, 2015, p. 26). The fourth is Domestication of Technology, which "continues where the social shaping of technology leaves off" (Baym, 2015, p. 52). This theory focuses its attention on how new technologies become integrated into everyday life and transition from being seen as a new technology into something that is ordinary and mundane. According to SST, technology is not developed along a linear path (Williams & Edge, 1996). Rather, it is influenced by countless socio-economical factors and can branch out in any number of different directions. This theory applies to a full circuit of technological development, from

product creation, to finding a place in the market, to how end users use and engage with the piece of technology. This research project focuses on the latter point, looking at how different pieces of technology (platforms) influence the ways users behave, and in a lesser fashion, how user behaviour influences those pieces of technology.

Social Shaping of Technology could be considered a piece of a modern evolution of Joshua Meyrowitz's (1986) classic research on the impacts of electronic media (mainly television) on social behaviour. In his time, television was the primary disruptor to traditional communication methods. He claims that new advances in digital media have the potential to create new and merge existing social spaces, reshaping standard social behaviours (Meyrowitz, 1986). Similarly, a primary tenet of SST is that new technologies will shape and influence the people who use them. This research project will illustrate this concept by studying user behaviour on three relatively established digital platforms.

Theory 1: Technological Determinism

For centuries, Technological Determinism has been a tool of opponents of new technology, used to decry new developments. The theory's base premise states that technology changes or influences users and society as a whole, usually for the worse. Presently, this could be applied to social media, digital platforms, and new advancements in artificial intelligence, among others. However, this is not a new theory. Socrates criticized the alphabet and the written word, foreseeing that people would lose their memories and intelligence and will "appear to be omniscient and will generally know nothing; they will be tiresome company, having the show of wisdom without the reality" (Baym, 2015, p. 28). In more modern examples, as the internet, and even the telephone, gained popularity, people voiced concerns about the alienating effects these inventions could have on offline social interactions. It is likely that every major invention in our

history, from the motor vehicle to electricity, dealt with opponents making the same claims.

Figure 2 below shows the general public's concern with new electric above-ground powerlines in the late 19th century.

Figure 1: "An Unrestrained Demon" from *Judge* Magazine, 1889

This illustration may seem comical to the modern viewer, but it illustrates the root causes of Technological Determinism: "our deep need to trust, connect with, and protect one another and ourselves, and the perpetual struggles these needs engender" (Baym, 2015, p. 44).

Theory 2: Social Construction of Technology

The opposing perspective to Technological Determinism is Social Construction of Technology (SCOT), which is based on the concept of users or society influencing the development of new technology. This theory "focuses on how technologies arise from social processes" in both their development and use (Baym, 2015, p. 44). According to SCOT, new technologies are not created in a vacuum; the social context in which they exist play an integral role in shaping advancement. The industrial revolution occurred because society demanded

faster, more efficient production. In the same way, new digital tools are developed and improved to meet the needs of society as a whole and individual users. Another perspective of SCOT is that users shape existing technology to best suit their needs. "The telegraph, radio, refrigerator, and internet are all technologies whose unexpected uses became their most common" (Baym, 2015, p. 46).

Theory 3: Social Shaping of Technology

If Technological Determinism is on one side of this spectrum, and SCOT is on the other, Social Shaping of Technology (SST) exists in between the two. SST recognizes that "people, technologies, and institutions all have power to influence the development and subsequent use of technology" and that "machines can and do accelerate certain trends, cultural weaknesses, and fortify certain social structures while eroding others" (Baym, 2105, pp 51-52). This theory sees the relationships between technology, users, and society as cyclical, with each side influence and being influenced by the other. Within a research and literature review, Williams and Edge (1996) "conceive of SST as a `broad church', without any clear `orthodoxy'" (p. 34). The authors also state that the boundary between technical and social is not clear, and this may affect the longevity of this theory (Williams and Edge, 1996).

Theory 4: Domestication of Technology

The principles of Domestication of Technology are similar to those of Social Shaping of Technology. Both theories agree that the technology-society relationship is two-way and each has an influence on the other. Domestication of Technology is, in a sense, a continuation of SST. It does not attempt to explain the nature of the relationship, as the first three theories do; instead, it describes the way new technologies are eventually integrated into our everyday lives and become mundane. The earlier example of above-ground power lines illustrates how over time,

technology that once caused controversy will become nearly invisible to critics as it is domesticated.

The Significant Social Impacts of Anonymity

The first section of this literature and policy review looks at the social effects caused or exacerbated by levels of anonymity. In the broad context of communications theory, this discussion relates to Habermas's definition of the public sphere. Habermas writes that: "Citizens behave as a public body when they confer in an unrestricted fashion--that is, with the guarantee of freedom of assembly and association and the freedom to express and publish their opinions-about matters of general interest" (Habermas, 1974, p. 49). In her development of public sphere theory, Fraser (1990) says that the idea of the public sphere is essential for open access, participatory parity, and social equality. As I explore in detail below, in the digital world, a truly realized public sphere would place everyone on equal ground. Digital platforms, such as social media, have created a new means for people to communicate and form public opinion. However, existing research points to the ways that people act in these situations are influenced by their degrees of anonymity – a relationship that has both antisocial and prosocial outcomes.

To explore these issues, I organize this literature into offline/online and prosocial/antisocial themes. The literature within these categories supports the argument that a heightened degree of anonymity online leads to more frequent toxic behaviour. Each of the authors cited in this section have varying definitions of what constitutes toxic, harmful, flaming, or aggressive behaviour on the antisocial side, as well as what constitutes helpful, kind, or benign behaviour on the prosocial side. Concepts of anonymity are similarly differentiated in the literature. Prior to the digital era, anonymity was studied as an antecedent to deindividuation (a state in which a person is not seen as or does not feel like an individual) (Diener, 1979). Diener's

theory of deindividuation has become a foundational source of this body of literature, as well as in the fields of psychology and sociology. "According to this theory, the deindividuated person is blocked by environmental factors from becoming self-aware and is thus less likely to regulate his or her behavior in reference to personal and social standards" (Diener, 1979, p. 1161). In other words, a person who is deindividuated is more likely to react quickly to emotional stimuli without worrying about the consequences. Famed psychologist Philip Zimbardo hypothesized that more deindividuation would lead to greater levels of aggression, and his experiments found that this hypothesis was correct in a natural setting, but not in the laboratory (Zimbardo, 1969). Zimbardo's findings illustrate Orne's concept of "demand characteristics of the experimental situation", which describes the circumstances surrounding a knowing or aware experiment participant (Orne, 1962, p. 779). Ultimately, a person who knows they are in an experiment or are being observed will not act according to their true nature. This is commonly referred to as the Observer Effect. Silke (2003) built upon both Diener's and Zimbardo's theories in a study of violent behaviour in Northern Ireland. By analyzing police reports, media stories, and materials from a victim support group, the author was able to code 500 different assaults in categories such as extremity of aggression and whether the attacker(s) was/were disguised. The results of this study show that in more violent assaults, the attacker is more often disguised, implying that people seek anonymity to conceal themselves when they perform heinous acts (Silke, 2003, pp. 496-497).

On the other hand, anonymity can provide security to individuals in different ways; as Jordan (2019) writes:

"Creative security, the right to experiment freely. Choice security, the right to be wrong, to make mistakes, particularly in artistic and creative practices. Personal security, the

right to address wrongs when you cannot deal with the consequences of addressing those wrongs. Judgmental security, the right to be judged on the criteria for the judgement and not on extraneous factors such as gender, race and appearance" (p. 575).

Anonymity can also protect vulnerable populations from threats based on their personal attributes, protect people's ability to peacefully protest without fear of losing employment, and allow people to freely express their opinions and beliefs without fearing the consequences.

As is the case with offline anonymity, existing research has found that there are prosocial effects of online anonymity. Suler's (2004) theory of the online disinhibition effect describes benign disinhibition as the prosocial side of what happens when people feel less restrained and express themselves more freely: "They reveal secret emotions, fears, wishes. They show unusual acts of kindness and generosity, sometimes going out of their way to help others" (p. 321). Anonymity is one of the core factors that creates the disinhibition effect. Lapidot-Lefler and Barak (2015) study the benign disinhibition effect in an experimental study; their findings show that increased anonymity and invisibility online can lead to more open expression and conversation (Lapidot-Lefler & Barak, 2015, p. 11). Users of digital platforms can create anonymous or pseudonymous accounts to safely follow accounts or people that their social circles or jobs might look down on them for following. It allows people the freedom to explore new areas and express themselves without fear of judgement. For example, an Indonesian study found that participants were more likely to participate in political discourse anonymously for the following reasons: "avoidance of conflict with their closest friends due to political differences, protection from online threats, avoiding embarrassment, as well as avoiding professional consequences" (Perbawani, 2018, p. 199).

However, the prosocial social effects of online anonymity are not reflective of the complete situation. Antisocial social effects are studied much more frequently, and many authors opt to include both within the same article (Chang, 2008; Clark-Gordon et al., 2019; Jordan, 2019; Kim et. al., 2019; Mikal et al., 2016; Omernick & Sood, 2013; Santana, 2014; Siegfried-Spellar & Lankford, 2018; Williams, 2006). Much of the literature in this field studies some aspect of how anonymity influences harmful behaviour online. This includes behaviour such as cyberbullying, aggression, hostile language, threats, violence, illegal activity, and sexually inappropriate content, among others. Based on the literature, these behaviours tend to occur more frequently or more severely when users are anonymous, deindividuated, or disinhibited. For example, in a longitudinal study, Wright (2013) found that anonymity related positively to cyber-aggression in young adults, and that participants were much more likely to post harmful content online when anonymous, because they believe they would not get caught and content online is not permanent. Additionally, people are more likely to share any content anonymously rather than identifiably, if given the choice, and much more likely to share controversial content anonymously (Zhang & Kizilcec, 2014). As Sparby (2017) writes:

> "When users are allowed to post anonymously, they are more likely to engage in cyberbullying or trolling because such behaviors cannot be linked to their real-world identities; they can say and do as they please while experiencing virtually no repercussions" (p. 86).

Suler's online disinhibition effect, discussed previously, describes toxic disinhibition as the antisocial side of what happens when people feel less restrained and express themselves more freely online: "rude language, harsh criticisms, anger, hatred, even threats. Or people visit the dark underworld of the Internet—places of pornography, crime, and violence—territory they

would never explore in the real world" (Suler, 2004, p. 321). This online disinhibition effect is one of the primary theories used to describe anonymous online interactions. Cyberbullying, in particular, is a common area of study, which may be because of increased mainstream media coverage of its real-world effects in recent years. Cyberbullying affects younger people (children, teens) more than adults, with often fatal consequences (CTV, 2012; Dean, 2012; Lowry, 2016). Wachs (2019) found that the perpetuation of cyberbullying was correlated with more severe cyberhate activities and that individuals who are involved with one form of aggression online are more likely to be involved with others. Therefore, anything that can be done by governments or platforms to limit online harms, even slightly, could have significant impacts in protecting users. That being said, a hallmark of democratic countries is that their governments do not regulate the free speech of their citizens, creating a challenging situation where they must balance these two competing priorities.

It is also important to note that different perspectives exist within the literature on the effects of anonymity. For example, some studies have found no correlation between higher levels of anonymity and antisocial online behaviour (Rosner & Kramer, 2016). However, the majority of literature surveyed found that removing or reducing anonymity does lead to more civility in online communication (Cho et al., 2012; Kurek et al., 2019; Moore et al., 2012; Santana, 2014; Zimmerman & Ybarra, 2014; Wu et al., 2017). This generally accepted finding influences the basis of this research project. If less anonymity has been shown to increase prosocial behaviour, and more anonymity increases antisocial behaviour, then further research is warranted into how specific digital platforms operate within the context of these assumptions.

Policy Review: Commercial Digital Platforms, Canadian, and International Governments

      This section provides an overview of the current digital environment experienced by Canadian users. This includes a review of two main areas: current Canadian policy within multiple government divisions and agencies; and the structures and policies of three popular platforms: Facebook, Reddit, and 4chan. This section also examines what other nations are doing in this space and summarizes authors' suggestions for potential policies and regulations.

Part 1: Design/policy Considerations of Commercial Digital Platforms

      When studying behaviour online, a researcher's first instinct may be to look at the content generated: what people post, how they react to others' content, what is shared, etc. However, the structure, policy, and design characteristics of online platforms are just as important in understanding why users behave a certain way. "How technical infrastructure is designed and administered is not only a technically complex function but one with significant public interest implications" (DeNardis, 2015, p. 761). This perspective relates to various socio-technical theories that are explored further within the last section of this literature review.

      The first platform included in this research is Facebook. Following its terms of service, a Facebook user must: "Use the same name that you use in everyday life. Provide accurate information about yourself. Create only one account (your own) and use your timeline for personal purposes" (Facebook Terms of Service). Based on these guidelines for user identification, Facebook is a platform with minimal user anonymity, and the least anonymous of the three being studied. To identify profiles that are fake, or not associated with a real person, Facebook employs advanced algorithms that have been shown to quickly learn signs of fraudulent profile creation (Pourghomi et al., 2020). Looking at content policies, Facebook does not allow hate speech, violent or graphic content, adult nudity, sexual activity, sexual

solicitation, or cruel and insensitive content (Facebook Community Standards). "Our current definition of hate speech is anything that directly attacks people based on what are known as their 'protected characteristics' — race, ethnicity, national origin, religious affiliation, sexual orientation, sex, gender, gender identity, or serious disability or disease" (Allan, 2017). In a 2018 statement, Mark Zuckerberg discussed the platform's strategy for proactively identifying harmful content using artificial intelligence (AI): "Facebook's AI is trained to identify terrorist content, self-harm, nudity, hate-speech, and fake accounts that spread misinformation" (Zuckerberg, 2018). This proactive approach to content moderation helps protect human moderators from the "psychological toll of having to look at the worst the Internet has to offer" (Gillespie, 2018, p. 123).

The second platform studied, Reddit, launched as a news aggregation service and has since become a popular social media platform for users to post original content, share existing content, link off-site, and comment on each others' posts. The site uses a mixture of employee and volunteer moderators to ensure guidelines are being adhered to (Moderator Guidelines for Healthy Communities). Gibson (2019) studied how different moderation policies shape self-censorship and content posted on Reddit. Her content analysis looked at two subreddits (individual communities on the platform with different moderation policies, of which there are thousands). Her findings showed that moderation policies "effectively set norms around style, affect and topic" (Gibson, 2019, p. 12). This statement supports that idea that well executed platform moderation is an important step towards reducing online harms. Reddit users are able to choose any username they want, giving the platform a semi-anonymous, or pseudonymous, structure. While a user's identity may be hidden, every action they take on the site is recorded to their profile, allowing other users to follow their path through the platform and see what they

have posted or commented on. While this could potentially allow a user to recognize or identify another based on content they post, Reddit takes a strong stance against revealing personal information of any user (Reddit Content Policy). This policy protects users who may not want their online activity traced back to their offline lives, for various reasons. One high-profile example of this policy in action has been the GameStop stock phenomenon. Late in 2020 through 2021, users of a sock market subreddit began researching and discussing GameStop as an interesting case study of hedge fund market manipulation. Users shared information on the platform, which ultimately led to an incredible increase of investors buying stocks in the company. As part of this occurrence, Reddit users were accused of market manipulation, and one user was even sued for securities fraud (Clark, 2021). Maintaining user pseudonymity for other users discussing GameStop became top priority.

On the far end of the anonymity spectrum is 4chan, an anonymous message board, which has no user registration process or identification requirements. All content and comments posted on a board are identified only by a randomly assigned post number, with no attribution to the poster themselves. "Posts are fully anonymous by default and very rarely contain pseudonyms or other identity signals" (Bernstein et al., 2011, p. 50). Dozens of individual boards exist that are dedicated to different content, ranging from the mundane (business and finance) to the socially taboo (pornographic requests). The platform's rules are not as formal or strict as Facebook or Reddit. They focus heavily on keeping specific content within specific boards and respecting the power of administrators and moderators, rather than dictating what users post or how they identify themselves (4chan Rules). The platform was an early home of Anonymous, the hacktivist group that has been responsible for online pranks, protests, and disruptions of the status quo (Reagle, 2015, p. 107). The structure of this platform creates an open playground for

behaviour not permitted in more mainstream spaces. "4chan's key interface design elements—anonymity, ephemerality, simple rules, and low moderation—help develop [its] ethos of lawlessness and unaccountability" (Sparby, 2017, p. 88).

The three digital platforms identified for research have very different structures, policies, and levels of user anonymity (see Figure 1 below). This research project intends to show how these differences, particularly the varying degrees of anonymity, influence user behaviour.
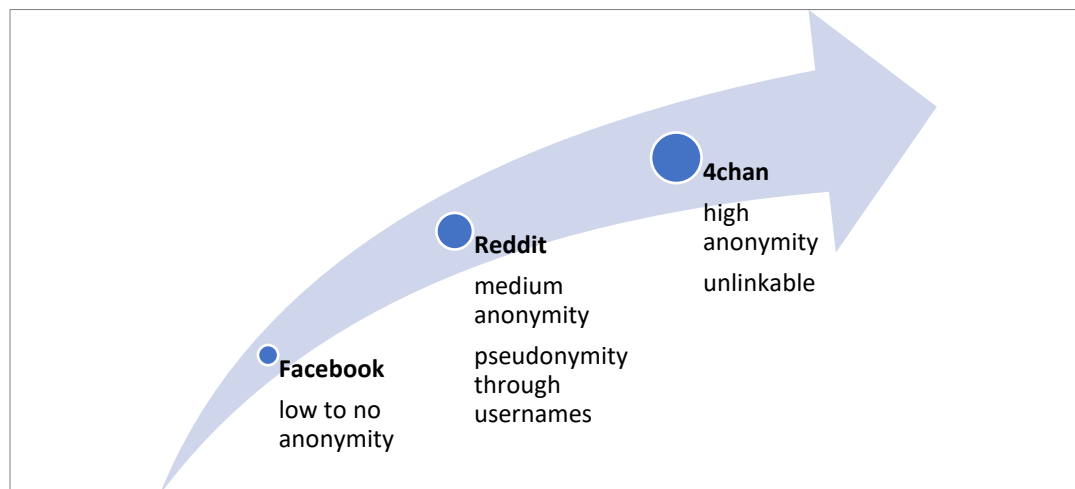
Figure 2: Anonymity Spectrum of Facebook, Reddit, and 4chan

Part 2: Canadian Communication Policy/regulation Regarding Online Anonymity

Much of the existing research on Canadian communication policy covers issues such as improved access for all Canadians and the preservation of culture through the Canadian Radio-television and Telecommunications Commission (CRTC) (Abramson & Raboy, 1999; Geist, 2016; Hackett & Anderson, 2011). At the time of this research, Canadian communication policies focus primarily on issues like privacy protections, data retention, and cybersecurity. However, they do not address concerns around harmful content or toxic behaviour online. One reason for this could be a reluctance of the federal government to foray into citizen censorship, as Section 2 of the Canadian Charter of Rights and Freedoms protects freedom of expression: "Everyone has the following fundamental freedoms: freedom of thought, belief, opinion and

expression, including freedom of the press and other media of communication" (Canadian Charter of Rights and Freedoms).

There are some public departments and agencies that are involved in regulating and policing different aspects of internet governance. The Office of the Privacy Commissioner of Canada deals with issues around protecting personal information, data, metadata, and security tips (Office of the Privacy Commissioner of Canada). Organizations like the Canadian Centre for Child Protection exist to protect children online in areas like harassment, online luring, and exposure to sexually explicit material (Canadian Centre for Child Protection). Bill C-13, which was enacted in 2015, adds a new offence to the Criminal Code against the "non-consensual distribution of intimate images" (Bill C-13 [Historical]). The Canadian Centre for Cyber Security helps organizations and businesses deal with cyber threats such as hacks, DDOS attacks and infrastructure vulnerabilities (Canadian Centre for Cyber Security).

Although there are not currently any federal regulations around online anonymity and protecting citizens from toxic behaviour, there may be in the near future. Canada's Digital Charter was created in 2019 as a guiding document for policy development, and was informed by public consultations that began in 2016. This plan highlights ways the government will advance Canada's digital leadership in the world by taking an "ambitious, aspirational, principled approach to digital and data transformation" (Canada's Digital Charter in Action, p. i). The Royal Canadian Mounted Police have recently (in 2015) created a digital policing strategy, known as Connected RCMP (Royal Canadian Mounted Police). This strategy outlines digital threats to Canadians, including "cyberviolence, such as use of social media and the Internet to incite hate; to bully or harass individuals" (Royal Canadian Mounted Police, p. 9). Bill C-11, which had its first reading on November 17, 2020, will enact the Consumer Privacy Protection

Act, which will "support and promote electronic commerce by protecting personal information that is collected, used or disclosed in the course of commercial activities" (Bill C-11). The Department of Justice Canada is currently holding public consultations on modernizing the Privacy Act (Department of Justice).

Despite these activities, at the time of this Capstone project there was not anything concrete in the works for how the government plans to address the regulation of digital platforms, other than stating that "social media platforms are already subject to the same laws as other organizations operating in the Canadian marketplace" (Fact Sheet: Digital Charter Implementation Act, 2020). Since the time of this research, Bill C-10 was introduced, which would grant the federal government more control over internet governance (Raman-Wilms & Curry, 2021). This Bill is mentioned in the conclusion of this project, but is not included within this literature review. However, as has been seen in multiple other instances, digital platforms often operate outside normal laws because they operate virtually in multiple sovereign nations, without having a physical presence, leading to convoluted legal cases as seen in Spain and India, among many other countries (Ribeiro, 2015; Phys.org, 2018).

In 2013, the Canadian government attempted to introduce Bill C-30, which would, among other things, amend items the Criminal Code "relating to hate propaganda and its communication over the Internet, false information, indecent communications, harassing communications, devices used to obtain telecommunication services without payment and devices used to obtain the unauthorized use of computer systems or to commit mischief" (Bill C-30 [Historical]) This Bill was met with public outcry against overstepping government control from both sides of the political spectrum and was not passed as law (Geist, 2016).

Canadian Government Policy Review	
Organization, Policy, Bill, or Report	**Purpose**
The Canadian Radio-television and Telecommunications Commission	Preserves Canadian culture/content, advocates for equal access to internet for all Canadians
Office of the Privacy Commissioner of Canada	Protects personal information, data, metadata, and user security
Canadian Centre for Child Protection	Protect minors from harassment, luring, sexually explicit material
Canadian Centre for Cyber Security	Helps organizations defend against as hacks, DDOS attacks and infrastructure vulnerabilities
Royal Canadian Mounted Police, *Connected RCMP*	Protect Canadians from cybersecurity threats, cyberviolence
Department of Justice Canada, *Privacy Act*	Protect Canadians' privacy, offline and online
Canada's Digital Charter	Advance Canada's digital leadership, literacy, and access
Bill C-11	Protect users' personal data online
Bill C-13	Criminalizes the non-consensual distribution of intimate images
Bill C-30	Reduce false information, harassment, hacking
Bill C-10 (future)	More control over internet governance (online content)

Table 2: Canadian Government Policy Review

The current Canadian regulatory landscape includes many initiatives to protect citizens' privacy, personal data, and cybersecurity, but does not discuss harmful content or toxic behaviour online. Some experts in the field say that Canada is falling behind in the global context of understanding and advancing mediated communication (Savage, 2008). It wasn't until 2021 with the introduction of Bill C-10, previously mentioned, that Canada has taken any tangible action towards this end. The following sub-section of the literature review looks at examples from other nations and highlights other approaches to internet governance.

Part 3: International Policy

To provide international context for this literature review, articles that analyze digital policies from other nations have been included. These include South Korea (Cho et al., 2012; Cho & Kwon, 2015), the United Kingdom (Horsman, 2016), Slovenia (Ceferin & Meznar, 2014), and Brazil, India, and China (Mahrenbach & Mayer, 2020). The governments of Brazil,

India, and China can also be studied to offer different policy options that the Canadian government could adopt. The study of these three governments is relevant to a North American audience because, according to World Bank, they are three of the "top four countries in terms of internet populations" (Mahrenbach & Mayer, 2020, p. 131).

One key issue is the ways that different countries organize the regulation of privacy protection. These seems to differ between national contexts. These approaches form the core debate in personal protection online: Is it the responsibility of the government to protect its citizens, or is it up to the platform or service provider to be a responsible corporation? For example, in Brazil and India, privacy protection seems to fall to the government, whereas in China it is the responsibility of the service providers (Mahrenbach & Mayer, 2020). In the United Kingdom, regulators are dealing with similar challenges as other democratic nations. Digital platforms hold the status of "'private communications service providers', remaining outside the confines of the Data Retention and Investigatory Powers Act 2014" (Horsman, 2016, p. 151). Privacy is a protected right in the United Kingdom, but laws still need to be upheld. The struggle here is which is deemed more important. "Services which seek to protect the identity of the individual have a profound impact upon … the ability of law enforcement officials to identify [and] prosecute for breaches of relevant legislation" (Horsman, 2016, p. 152).

One possible option to appease both sides of this discussion that has been used in some European countries (and Australia) would be changes in government regulation to force platforms to retain data and records for a set timeframe to assist with prosecution of illegal activity. Within the past five years, Germany enacted the Network Enforcement Act (2017) to help enforce the law on social media, Australia has the Enhancing Online Safety Act (2018), and France updated its Electoral Code (2018) to prevent the spread of misinformation online (Simona

et al., 2019). Under Slovenian law, individuals can be prosecuted for posting unlawful content online, but in many cases, anonymity protects these users. In Slovenia, individual user accountability is uncertain. Ceferin and Meznar (2014) argue that the platforms or publishers should be held accountable for anonymous hate speech posted on their media if a person or group cannot be identified for civil litigation. South Korea has enacted a Real Name Verification Law in 2007 for online platform use, which places more accountability on users for the content they post. Cho (2012, 2015) published two key articles that provide empirical evidence that decreasing anonymity online leads to more civil behaviour on those sites affected by the law. However, it is also worth noting that participation in online dialogue decreased after the law was put in place: "some potential commenters may withdraw themselves from writing comments due the inconvenience and risk" (Cho & Kwon, 2015, p. 370).

This sub-section of the literature review provides an international perspective on how elected governments can get involved with internet governance. Perhaps the Canadian government could glean insight from these case studies and blend them into a solution that works for Canadians.

Summary

This literature and policy review introduced this project's theoretical framework, and existing literature that I will build upon. I grouped the literature thematically into two main bodies (the social effects of anonymity and the current state of government policy and platform policy regarding issues of online anonymity). Whether studied offline or online, many authors have looked at how anonymity influences society. The findings within the literature are mixed but skew heavily towards agreeing that anonymity influences online posting behaviour in antisocial ways more than it does in prosocial ways. Governments and platforms have reacted to

this tension in various ways. Government policy must balance freedom of speech with the antisocial implications of online behaviours that can be shaped and shielded by degrees of anonymity. Platforms have also introduced ways to extend or limit degrees of anonymity, and moderate user behaviours in different ways. Overall, this section shows that both governments and platforms have roles to play in protecting users from online harms. Based on the review of the literature, I argue that governments have an obligation to intervene when corporations directly impact the lives of their citizens, and while platforms may describe themselves as intermediaries, not publishers, they should accept some level of responsibility for providing the space for harmful content to be created and shared. As Gillespie writes: "Platforms may not shape public discourse by themselves, but they do shape the shape of public discourse" (p. 23).

Based on the gaps identified through this chapter, this capstone project will explore the following research questions: "How do anonymity, policy, and structure affect social media users' behaviour online, and how does user behaviour shape platform structure?" As well, based on the literature reviewed, I developed the following hypothesis to guide my inquiry: Platform users who are more anonymous (as defined through platform rules and policies) are more aggressive than platform users who are less anonymous.

The next chapters develop my research design, which will look specifically at how the different user anonymity structures of three popular digital platforms (Facebook, Reddit, and 4chan) influence the content being posted online and what is currently being done at the government and platform levels to reduce online harms for users. To answer the research question and explore my hypothesis, I will perform a content analysis of existing public social media posts. The following chapter describes this project's research design and methodology, using a theoretical framework following the theory of Social Shaping of Technology.

Chapter 3: Research Design and Methodology

Introduction

In this chapter, I outline the strategies I use to conduct my research, including the study design, methodology, data collection, storage, and analysis, and coding procedures. The previous Literature Review chapter focuses on the content and findings of publications. However, the methods these authors used to collect data and reach their conclusions are also interesting and helpful for my research design. This first section of my chapter on methodology looks at the ways authors and researchers studying internet governance approach data collection and analysis. To inform my own study, I review this material and identify the most- and least-used methods to discover trends and best practices.

The existing literature features a balance of quantitative and qualitative research methods, as well as many studies that use a mixed-methods approach. Table 2 categorizes these sources by their primary methodology, though many may feature a second method as well. From this grouping process, it is apparent that content analysis is the most popular choice of methodology used by researchers studying this topic, followed closely by surveys/questionnaires, engaged scholarship, experimental design, and critical discourse analysis. At the tail end of this list (used significantly less often), are case studies, literature reviews, and interviews. It is logical that content analysis is the most used method to associate online anonymity with toxic behaviours, as it provides researchers with a direct look at users' activity on digital platforms. My methodology within this research project follows the majority consensus and employs a content analysis to observe natural user behaviour online.

Primary Methodologies in Existing Literature							
Content analysis	**Survey or questionnaire**	**Engaged scholarship**	**Experimental design**	**Critical discourse**	**Case study**	**Literature review**	**Interview or focus group**
Bernstein, M.S. et al. (2011)	Kasakowskij, R. et al. (2018)	Chang, J. (2008)	Deiner, E. (1979)	Abramson, B. D., & Raboy, M. (1999).	Ascher, D. L. & Noble, S. U. (2019)	Clark-Gordon, C.V. et al. (2019)	Milosevic, T. (2016)
Cho, D. et al. (2012)	Kim, K.K. et al. (2019)	DeNardis, L. & Hackl, A. M. (2015)	Lapidot-Lefler, N. & Barak, A. (2012)	Ceferin, R., & Meznar, S. (2014)	Grimmelmann, J. (2015)		
Cho, D., & Kwon, K. H. (2015)	Kurek, A. et al. (2019)	Geist, M. (2016)	Lapidot-Lefler, N. & Barak, A. (2015)	Fraser, N. (1990)	Mikal, J. P. et al. (2016)		
Gibson, A. (2019)	Lang, G., Subramanian, R., & Mavlanova, T. (2014)	Gillespie, T. (2018)	Lea, M. et al. (2001)	Habermas, J. et al. (1974)			
Mahrenbach, L. C., & Mayer, K. (2020)	Lowry, P. B. et al. (2016)	Hackett, R. A., & Anderson, S. (2011)	Orne, M. T. (1962)	Jordan, T. (2019)			
Moore, M.J. et al. (2012)	Perbawani, P. S. et al. (2018)	Horsman, G. (2016)	Pourghomi, P. et al. (2020)	Moll, M., & Shade, L. R. (2011)			
Omernick, E., & Sood, S. O. (2013)	Seigfried-Spellar, K. C., & Lankford, C. M. (2018).	Jane, E. A. (2015)	Rösner, L., & Krämer, N. C. (2016)	Reagle, J. M. (2015)			
Santana, A.D. (2013)	Wachs, S., et al. (2019)	Savage, P. (2008)	Zhang, K.& Kizilcec, R. (2014).	Pfitzmann, A., & Köhntopp, M. (2001)			
Silke, A. (2003)	Wright, M.F. (2013)	Suler, J. (2004)	Zimbardo, P. G. (1969).	Simona, M. et al. (2019)			
Sparby, E. M. (2017)	Wright, M.F. et al. (2019)	Williams, K.S. (2006)	Zimmerman, A. G., & Ybarra, G. J. (2016)				
Suh, K.-S. et al. (2018)	Wu, S. et al. (2017)						
Young, R. et al. (2018)							

Table 3: Analysis of Methodologies Used in Literature

Generally, researchers in this literature employed content analyses by analyzing user-generated content on one or more digital platforms to understand some facet of user behaviour, platform structure, or government policy. For example, Bernstein et al. (2011) studied 5,576,096 individual posts on 4chan to study the ephemeral nature of the platform's structure. Cho et al. (2012) and Cho and Kwon (2015) compared online user behaviour pre- and post-implementation of South Korea's Real Name Verification Law, which removed user anonymity, by analyzing the content of comments posted on popular news sites. Santana (2013) sampled data from two different platforms (in this case newspaper sites), one of which allowed anonymity and one of which did not. He compares user behaviour between the two sources, while maintaining one common subject matter. Santana's approach informs my research design in this project.

The following sections of this chapter provide a more detailed overview of the theoretical framework used, the study design, data collection, data storage, and data analysis.

Research Methods: Study Design

According to DeNardis et al. (2020), there is a need for "scholarship that makes visible to society the sinews of power constructing and controlling the Internet and explaining what the implications are for society and the economy" (p. 2). Inspired by this approach, this research project aims to contribute to efforts to conduct this kind of research through a content analysis of first-level comments (branch comments) and replies to these comments (leaf comments) on an initial, public post (root post) on three digital platforms with different structures. To compare comments more accurately across platforms, I selected a single subject that was discussed on all three platforms. This analysis will look at the Canadian government's May 2020 decision to ban specific types of firearms. The content sample will be selected from the first week of May.

Research will be focused on analyzing comments on a post, rather than a post itself. The source of content implies different intentions; a user who shares harmful or aggressive content might have less antisocial intent than a user who generates their own harmful comment (Kim et al., 2019). Therefore, I only analyzed text-based, user-generated responses, rather than image or gif replies. For Facebook and Reddit, the initial post could be any image, text, or video post. On 4chan, any thread on a message board would be appropriate to gather material from. There has been research done on the influence of seed comments (replies to posts that spur further threaded discussion), and this will be used to help structure data collection and storage (Suh et al., 2018; Young et al., 2018).

In order to analyze a meaningful sample that contained lively conversation and had the potential to be polarizing, I decided to choose a subject matter to gather data from that could be seen as contentious. On May 1, 2020, the Government of Canada announced a new ban on assault-style weapons, which was met with mixed reactions. Proponents lauded the decision as the right choice to keep Canada safe, while opponents saw it as an infringement of their rights. I chose this topic because I assumed that since people traditionally held strong beliefs on this issue, opposing points of view would generate heated online discussion, and therefore provide a dataset worth analyzing. I anticipated that a more benign discussion subject like soap operas (Baym, 2015) or bass guitar (Jurich, 2019) would be less controversial and more prosocial, and therefore not conducive to the goals of my study.

To organize my data, I adopt the metaphor of a tree to represent different types of content presented on the three online platforms (Suh et al., 2018). The initial post can be viewed as the roots of a content tree. Each reply to an original post is a branch, and each additional reply on these branches are leaves.[1] As discussed in more detail below, in this project I collected and analyzed data from both branches and leaves, with the root post functioning as the source material that generates all relevant data. Some of these decisions are related to the structure of the platform posts: for Facebook and Reddit, the initial root post could be any image, text, or video post. On 4chan, any thread on a message board would be appropriate to gather material from.

Each of the three platforms studied in this project has a different technical structure, which presents users different queues about content. Facebook (Figure 3) highlights the root post at the top of the user's screen, followed by branch and leaf posts beneath it, organized either chronologically, by relevance, or by popularity. Leaf posts are clearly identified by a visual indent, nesting them beneath the branch that spawned them. Reddit (Figure 4) follows a very similar visual structure, with the root post highlighted at the top, followed by branches and nested leaves as you scroll down the content.

---

[1] In existing research, these branches have also been called seed comments (i.e., replies to posts that spur further threaded discussion) (Suh et al., 2018).

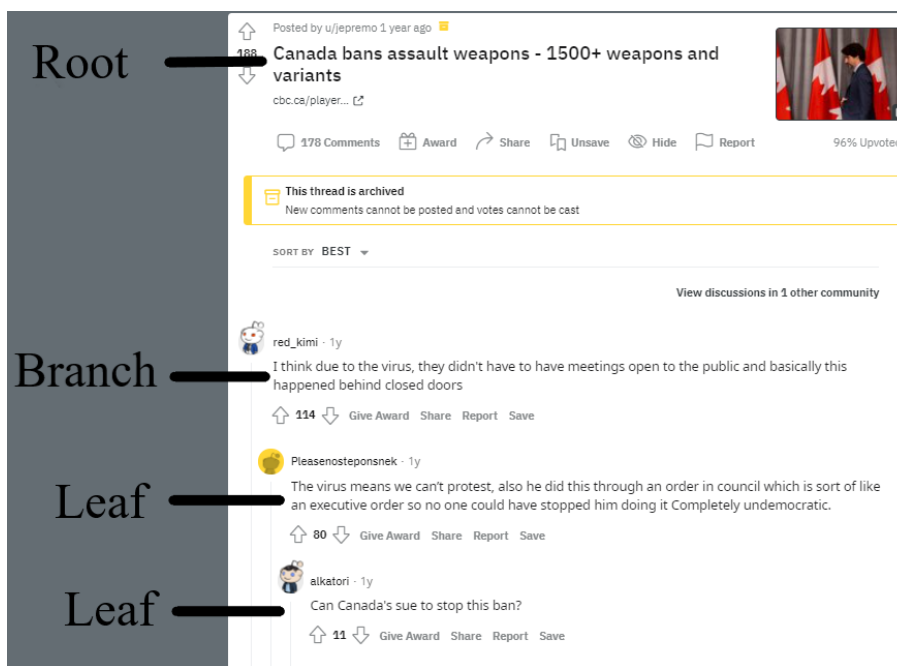Figure 3: Facebook Content Hierarchy

Figure 4: Reddit Content Hierarchy

On 4chan, however, users see a slightly different structure. There is no clear indentation to separate branches from leaves. Instead, replies to comments (leaves) are identified with a numbered code that appears at the beginning of that post (Figure 5). During the coding process of this research project, extra attention was given to the 4chan data to ensure each comment was properly coded according to the hierarchy.

Figure 5: 4chan Content Hierarchy

My dataset consists of comments (or replies) on a single post, rather than on a series of posts. I decided to focus on responses to posts because the source of content implies different intentions on the part of users; for example, a user who shares a harmful picture generated by someone else might have less harmful intent than a user who generates and posts their own harmful comment (Kim et al., 2019). To simplify the coding and analysis processes I restricted my data to textual comments, omitting images, videos, and gif replies. I selected a total sample size of 100 comments per platform, as this should be sufficient to obtain generalizable results. I note that "overall, users are more likely to consume media, both anonymously and non-

anonymously, rather than produce content" (Kasakowskij, 2018, p. 34). Therefore, no volume of content analyzed will be completely representative of the platform audience. I collected data with a non-probability, exploratory sample using a mix of convenience and purposive tactics. An exploratory approach was chosen because it "is used as a way of probing relatively unexplored topics and as a route to the discovery of new ideas or theories" (Denscombe, 2017, p. 24). A non-probability sample was more appropriate than a random sample due to the nature of this research design. Posts about specific subject matter (gun control) were sought out and selected. User demographics do not come into play within this research design, although the profile name from Facebook, username from Reddit, and post identification number from 4chan were recorded. This study also looked for root posts with multiple voices and many participants to avoid a situation where a handful of users took over the conversation with sub-conversations.

All research was conducted in Edmonton, Alberta, Canada, between April and June 2021, under the supervision of Dr. Rob McMahon in fulfilment of Capstone Project requirements for the University of Alberta's Master of Arts in Communications and Technology program. The data sample was limited to online posts made on May 1, 2020.

Research Methods: Data Collection and Storage

Within this project, data was gathered by searching public posts on Facebook and Reddit, for mentions of the May 1, 2020 Canadian firearm ban. A specific time window was selected for these root posts: only posts from May 1 were considered. By narrowing the timeframe of the root post, the content analyzed is closely comparable across all three platforms. I made the decision to keep this variable constant to partially mitigate the risk of users developing different perceptions based on time exposed to the subject matter, which could potentially influence the content they post online.

To collect the sample, I searched for specific terms on each platform: "Canada firearm ban." Facebook and Reddit both store public posts indefinitely, with Reddit archiving posts after six months to conserve storage space and prevent further commenting. This structure allows easy access for searching past content. However, since 4chan's structure does not store historical data that far in the past, a third-party website was used. Because 4chan's structure is based on impermanent ephemera, they do not store historical posts. Therefore, I used one of the many user-made archive sites for 4chan content, desuarchive.org.[2] This is an interesting example of Social Shaping of Technology in action; users clearly have a desire to access previous content and have created their own platforms to do so. Perhaps we will see 4chan implement a feature such as this in the future. Additionally, this structural element may have an influence on the types of content users post. If it is known that content disappears fairly quickly, users may be encouraged to be more aggressive or post more socially inacceptable content.

The same keywords and time window were used for all three platforms. These keywords were sufficient to produce results on each platform that possessed enough branch and leaf comments for the purpose of this study. The original plan was to find one initial post on each platform with approximately 100 branch comments. As I built out my methodology and identified a difference between branch and leaf comments, I decided to separate the two and collect a total of 100 comments from either source. Table 3 provides direct links to the posts that were used for this research.

---

[2] In earlier iterations of this plan, a different service called yuki.la was going to be used. However, that website's server went offline around the time this research was about to begin, so the researcher chose to amend the plan rather than waiting for it to come back online.

Selected Posts		
Facebook Post	Reddit Post	4chan Post

Table 4: Selected Root Posts

Once data was identified, the webpages were saved as .pdf files, which maintained the contextual integrity of the content, including the hierarchy of root posts, branch comments, and leaf comments. For visualizations of these layouts, see Figures 3-5 above on pages 38-39.

Following this phase of data collection, all content was imported into an open-source digital qualitative content analysis tool, Taguette, which will be described in more detail in the following Data Analysis section. I follow University of Alberta policies and Research Ethics Board guidance surrounding data management. Since this data is not being shared with other individuals or groups, there was no need to use a secure environment, such as Dataverse (provided by the University), and because this is all publicly accessible information, storage encryption was not necessary. The saved screenshots and accompanying text file were stored on a password-protected laptop, with a backup copy saved on a USB drive, kept in a personal safe, and only known about by the primary researcher, maintaining user anonymity and privacy.

Research Methods: Data Analysis

My intent with this research is to see how the level of anonymity afforded by each platform (Figure 6) may influence the content being posted by users. To achieve this, a number of technical steps must be followed, such as appropriate coding and analysis methods.
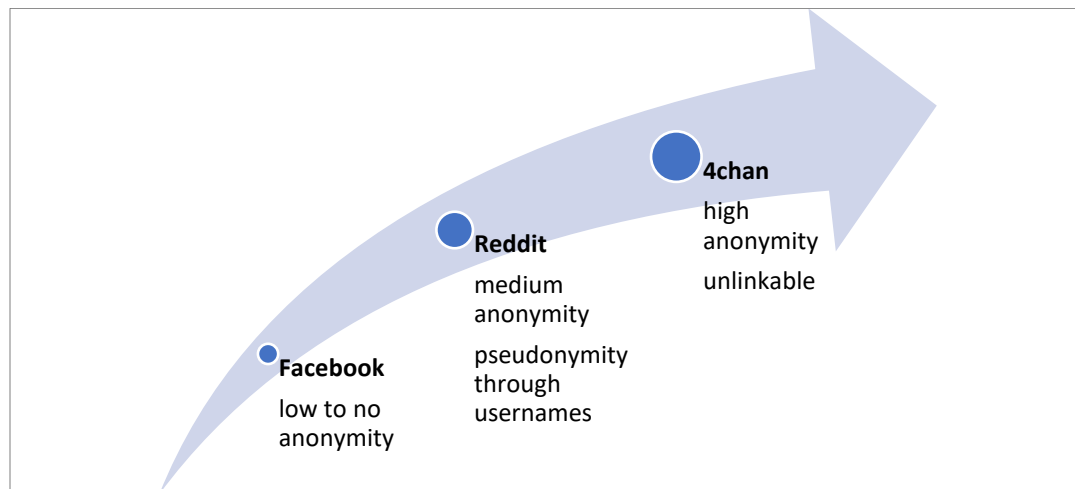

Figure 6: Anonymity Spectrum of Facebook, Reddit, and 4chan

One of the most important defining characteristics of a qualitative content analysis is the coding system employed by the researcher(s). In this study, I used a combination of deductive and inductive approaches to coding. I believe that taking a mixed approach allows for stronger refinement and results. Prior to data analysis, I compiled a list of preliminary themes that formed the basis of the final coding list used. Saldaña (2013) lists 32 different coding methods, that can be used independently or concurrently, which formed the methodological basis of this study's coding process (pp. 261-268). He also states that coding is a cyclical process with specific methods more appropriate for first or second cycles (Saldaña, 2013, p. 58). Within this current research study, an initial coding list was created based on examples found in other literature and by following Saldaña's (2013) 32 methods of qualitative coding, followed by a more refined list of codes and categories that emerged from a first pass of the data. Of these 32, I used five types: descriptive, belief/values, emotion, hierarchy, and domain/taxonomy. I then tested this list

through an initial pass of the raw data, which resulted in edits and revisions, detailed in the

following paragraphs. This research project followed a simultaneous approach to coding, in

which two or more different codes are applied to a single datum (Saldaña, 2013, p. 267). This

initial list mainly included generic descriptive codes such as 'swearing' and belief/values codes

such as 'freedom'. The actual application of these codes was done through an open-source digital

tool called Taguette, which allows users to import data into its interface to facilitate manual

coding. Specific words or phrases can be highlighted, assigned codes or categories, and recalled

at a later time for further analysis. All data is stored within the user's own device, alleviating any

concerns about insecure data storage. An additional benefit of this system is that it allows

multiple users to code the same data simultaneously, though this researcher chose to be the only

coder. A complete list of codes used within this study can be found in the following chapter.

Coding Procedure

As a reminder, a single root post was selected from each of the three platforms on the

topic of the Canadian government's May 2020 firearm ban, and a total sample size of 100

comments was sought (mix of branches and leaves). Each branch post (reply) can have multiple

leaves (replies to replies) that become branches on their own. To avoid becoming too convoluted,

each level past the initial branch is simply referred to as a leaf. After an initial review of the data,

I determined that a limit should be placed on how many leaf comments to analyze for each

branch. This was due to a requirement to manage the Facebook data, which produced an

exceptionally high number of leaf comments from a single branch (e.g., one branch example

generated over 100 leaves). To address this, I set a limit of 15 leaves per branch across all three

platforms in order to capture a broader sample and avoid focusing too much on a single

conversation between users. The majority of branch posts are not affected by this design choice;

it simply allows for more variety in the content analyzed. Within the coding process, the root post of each platform's data source, shown below in Figures 7-9, were omitted from all qualitative coding. I did this because I view the root post as the source of data generation rather than an analyzable datum.
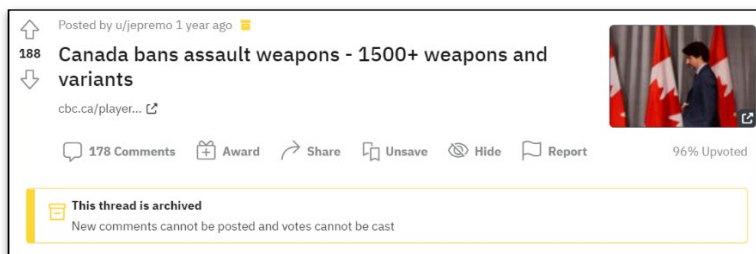

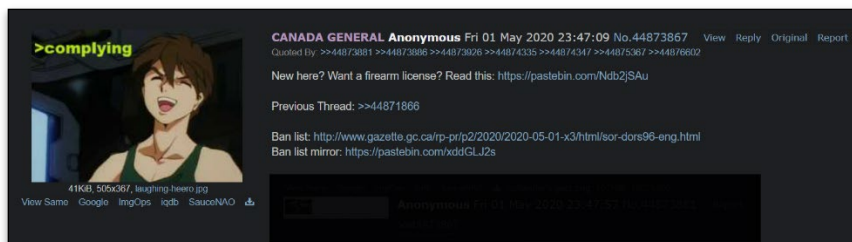Figure 7: Facebook Root Post

Figure 8: Reddit Root Post

Figure 9: 4chan Root Post

Of the five identified coding techniques (descriptive, belief/values, emotion, hierarchy, and domain/taxonomy), three were applied universally to all data: hierarchy, descriptive, and domain/taxonomy. A hierarchal code was assigned to each piece of content for data organization, but this was not used in analysis; rather labelling content as a root, branch, or leaf helped me organize the data and ensure I was following my sampling plan. All comments were assigned a descriptive code and domain/taxonomy code. According to Saldaña (2013), descriptive codes are used to broadly categorize data and "summarizes in a word or short phrase … the basic topic of a passage of qualitative data" (p. 88). This method of coding was used within my research to identify users' positioning on the subject matter: whether they were opposed to or supportive of the firearm ban, or if their comment was indifferent or unrelated to the subject. I used three different categories of domain/taxonomy codes; this type of coding is described by Saldaña (2013) as a way to "organiz[e] categories of meaning from participants" into subsections of a theme. I created three different themes, or domains, with multiple subsections, or taxonomies. These codes were used to identify aggressive or non-aggressive behaviour, and classify each post as a type of aggression or non-aggression for data analysis. The three domains used are "aggression on topic/in general", "aggression towards user", and "non-aggression". The taxonomies within these domains used include swearing, racism, threats of violence, asking/answering a question, and complimenting a user. Initially, I included an emotion code (anger, sorrow, fear, relief, joy, indifference) and belief/values code (safety and freedom) to generate supplemental data for analysis. However, after the initial coding, I decided to remove both of these categories because I deemed them irrelevant to the study and they did not often appear within the data. The resulting list was used as the final coding tool and is presented below in Table 5.

Category	Codes	Description
Hierarchal	Root	This category arranges comments by hierarchy. A root post is the original post; a branch is a reply or comment to the root post; a leaf is a reply or comment to the branch.
	Branch	
	Leaf	
Descriptive	Opposition to Ban	This category of codes looks at whether the content is in support, opposition, or indifferent to the ban, or if the content is unrelated.
	Support of Ban	
	Indifferent/Unclear	
	Unrelated	
Domain/Taxonomic (Aggressive in general or on topic)	General aggression	This category is aggressive content directed towards the topic, or just in general. There is no specified object to the aggression.
	Swearing	
	Gender/sexual slurs	
	Racism/slurs	
	Threats or violence	
Domain/Taxonomic (Aggressive towards another user)	General aggression	This category is aggressive content directed specifically towards another user of the platform.
	Swearing	
	Gender/sexual slurs	
	Racism/slurs	
	Threats or violence	
Domain/Taxonomic (Non-aggressive in general, on topic, or inter-user)	Benign comment	This category covers all non-aggressive content.
	Asking a question	
	Answering a question	
	Compliment to/about user	
	Compliment toward topic or in general	
	Defending a user	

Table 5: Final Coding List

With this coding list, I then went through all the posts selected for analysis and applied codes within Taguette, the open-source coding tool I used for this project Figures 10 and 11 below illustrate how a single 4chan post appears on the platform itself, followed by how it appears within Taguette. Figure 11 also shows how multiple different codes can be applied to a single post, following a simultaneous coding method, as described earlier in this chapter. The following codes are highlighted at the bottom of Figure 8:

- Descriptive-Opposition
- Hierarchal-Leaf
- Taxonomy-Aggression-General/Topic-Racism/Slurs
- Taxonomy-Aggression-General/Topic-Swearing
- Taxonomy-Aggression-General/Topic-Threats/Violence.
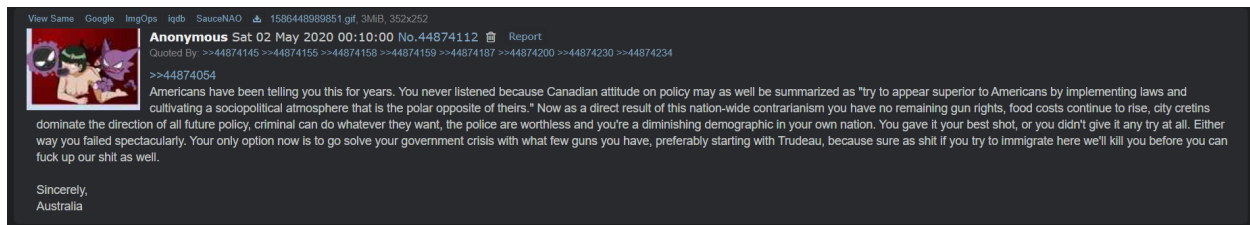
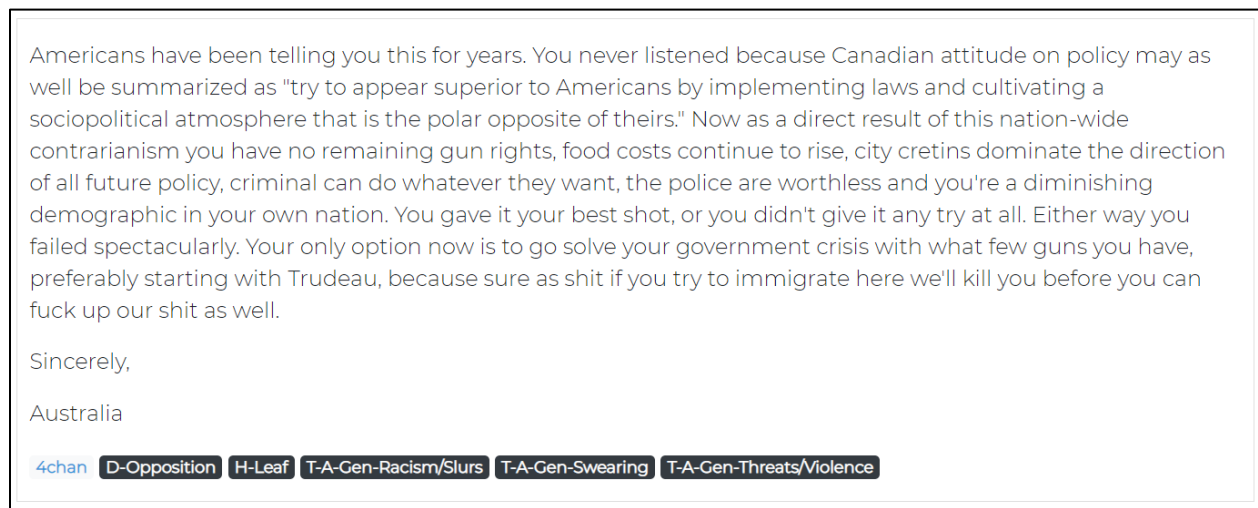Figure 10: 4chan Post Example

Figure 11: 4chan Post within Taguette

Summary

This chapter provided an overview of this study's research design and methodology. To answer this study's research questions, an understanding of user behaviour on varying platforms is needed. A coded content analysis was selected to be the most effective method of obtaining this data. Technological Determinism was considered as an option for the theoretical framework, but ultimately Social Shaping of Technology was chosen, as its cyclical nature fits better with this design and my personal perspective. Data will be collected and stored in accordance with University of Alberta policies. Data will be analyzed with the open-source software Taguette, following various coding methods presented by Saldaña (2013).

The following chapter presents the findings of the content analysis and offers a thorough discussion of their application to this study's research questions: How do anonymity, policy, and structure affect social media users' behaviour online, and how does user behaviour shape platform structure?

Chapter 4: Findings and Discussion

Introduction

In investigating the research questions and following the methodology and study design outlined in the previous chapter, I collected and analyzed data from my sample, which is explored within this chapter of my Capstone project. Based on my literature review, I expected that the more anonymity granted to users by specific platform structures and policies, the more likely they would be to post aggressive content. This assumption led me to hypothesize that the results of this study's content analysis would show that 4chan produced the highest volume of aggressive content, followed by Reddit, then Facebook, following each platform's level of user anonymity.

However, the results of this research project were not aligned with my preconceptions and expectations based on the existing literature that was reviewed. While the most anonymous platform, 4chan, did produce the highest number of aggressive comments, the least anonymous, Facebook, produced the next highest. My analysis of the data also shows that there is a difference in severity of aggression by platform, which was identified by assigning multiple codes to a single post if more than one aggression indicator was present in the content. Additionally, I discovered that each platform's users seem to prefer specific types of aggression, such as Facebook having more user-directed aggression and Reddit having more sexual/gender-based aggression.

This chapter explains the results of the data analysis with examples of content and concludes with a summary of this research's results before leading into the concluding chapter that includes study limitations and recommendations for future research.

Results of Data Analysis

User Positioning on Issue

As mentioned in the previous chapter on methodology, I used descriptive codes to identify users' positioning on the subject matter: whether they were opposed to or supportive of the firearm ban, or if their comment was indifferent or unrelated to the subject. I expected that there would be fairly equal representation of opponents and proponents of this topic. However, after data analysis, it became evident that the majority of comments were either indifferent or unrelated to the topic. The next most prominent set of comments were in opposition of the Canadian firearm ban and of gun control in general. With the exception of posts on Facebook, my data set contained far fewer supporters of the ban than expected. As is evident in the chart below (Figure 12), Facebook data showed a moderately balanced dispersion of comments that were in support, opposed, indifferent/unclear, or unrelated. In contrast, data from Reddit and 4chan showed highly uneven results of user positioning on this issue, with very few users being in support of the ban.

I suggest that this difference may be related to the technical characteristics of the platforms. On Reddit and 4chan, users seek out and join communities that relate to their interests and passions. These platforms meet some of the characteristics of an affinity spaces, which are "physical, virtual, or hybrid spaces of shared passion, organized around a common interest or activity" (Jurich, 2019). Characteristics of affinity spaces include "1. Common endeavour, not race, class, gender or disability, is primary" and "2. Newbies and masters and everyone else share common space" (Gee, 2005). Because there is a sense of group identity on Reddit and 4chan, users become deindividuated and "join a collective identity of users who have a specific, often strong, sense of themselves as a social unit (Sparby, 2017). In contrast, on Facebook, users

can come across this content more organically due to the structure and function of its News Feed. As there is no affinity space created on a public Facebook post, users are acting as individuals with their own opinions.

To examine user positioning on this issue in more detail, I applied four descriptive codes to the data. These codes indicate whether a comment was in opposition to the ban, in support of the ban, if the comment was indifferent or the position was unclear, or if the comment was completely unrelated to the topic. The chart below shows the results of these descriptive codes within each platform.
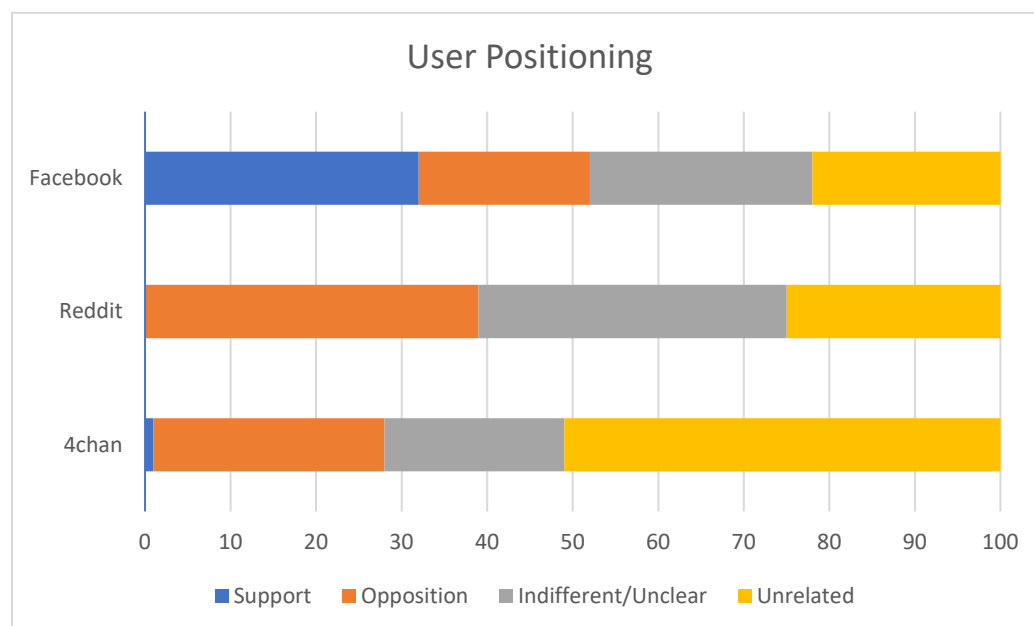

Figure 12: User Positioning

As Figure 12 illustrates, Reddit and 4chan both skew disproportionately towards opposition rather than support (39-0 and 27-1, respectively). That said, both of these platforms have a majority of posts that are either unrelated or indifferent/unclear. Facebook users have a fairly balanced dispersion of positions (32, 20, 26, 22). I believe that the different points of view present on Facebook are a major contributor to the high levels of user-directed aggression on this platform, which will be discussed further in this chapter.

Aggressive or Non-aggressive Behaviour

My analysis of the domain/taxonomy codes showed an interesting finding: all three platforms were fairly balanced in terms of aggressive user behaviour. As a reminder, the types of aggressive, or antisocial, behaviour I am looking at in this study are: general aggression, swearing, gender/sexual slurs, racism/slurs, and threats or violence. Of the 100 comments I analyzed on each platform, I labelled 29 4chan posts as aggressive, compared to 21 from Reddit and 25 from Facebook. The data show that 4chan had the highest number of aggressive posts, the highest number of aggressive codes, and the largest gap between the number of posts and the number of codes. Following Saldaña's (2013) method of simultaneous coding, some posts coded as aggressive were assigned multiple codes. For example, if there were instances of swearing and racism in a single post, it was assigned both codes. This is illustrated in Figure 13 below, where a single post is assigned two taxonomic codes.

> because sure as shit if you try to immigrate here we'll kill you before you can fuck up our shit as well.

Why would I want to move from one fucked up liberal managed chinese colony to another?

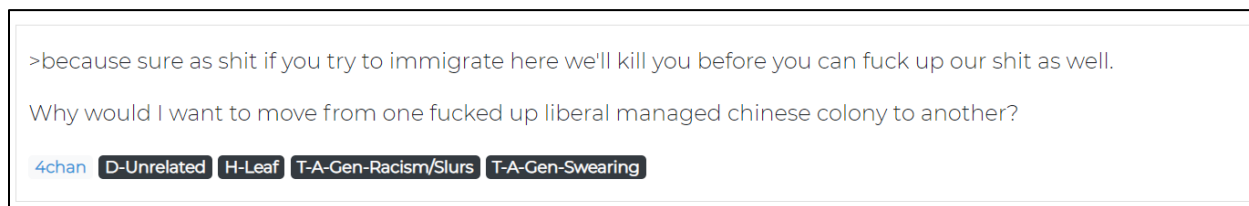4chan [D-Unrelated] [H-Leaf] [T-A-Gen-Racism/Slurs] [T-A-Gen-Swearing]

Figure 13: Example of Single Post with Multiple Aggression Codes

This gap between number of posts and number of codes can be attributed to the coding strategy I used (assigning multiple aggression codes to a single post) and expressed as a simple ratio. For example, Table 6 below highlights the level of aggression seen per post across all three platforms. 4chan showed a higher number of posts coded as aggressive (29), followed by Facebook (25) and Reddit (21). 4chan also showed the highest number of aggressive codes (43), followed by Facebook and Reddit (27 each). By reducing the ratio of aggressive posts to aggressive codes to a (1:x) figure, the relative level of aggression per post becomes apparent.

Therefore, Facebook's ratio of 25:27 becomes 1:1.08. Reddit's 21:27 becomes 1:1.29, and

4chan's 29:43 becomes 1:1.48.

Level of Aggression per Platform			
Platform	**Aggressive Posts**	**Aggressive Codes**	**Ratio (post:codes)**
Facebook	25	27	1:1.08
Reddit	21	27	1:1.29
4chan	29	43	1:1.48

Table 6: Level of Aggression per Platform

In line with my hypothesis and the findings of other research, the most anonymous

platform, 4chan, produced the most frequent and severe antisocial behaviour. However, based on

the literature I reviewed as part of this research project, I expected the results of this research to

show that the increased anonymity of users on 4chan influenced their behaviour much more

significantly, producing an even more disproportionate number of aggressive posts on 4chan

compared to the other two platforms.

Figures 14 and 15 on the following page provide a visual comparison of the volume and

severity of aggression and non-aggression across the three platforms. The gaps I discussed are

clearly present in the chart showcasing aggression. The two charts show that the majority of

posts on all three platforms are non-aggressive, or prosocial.
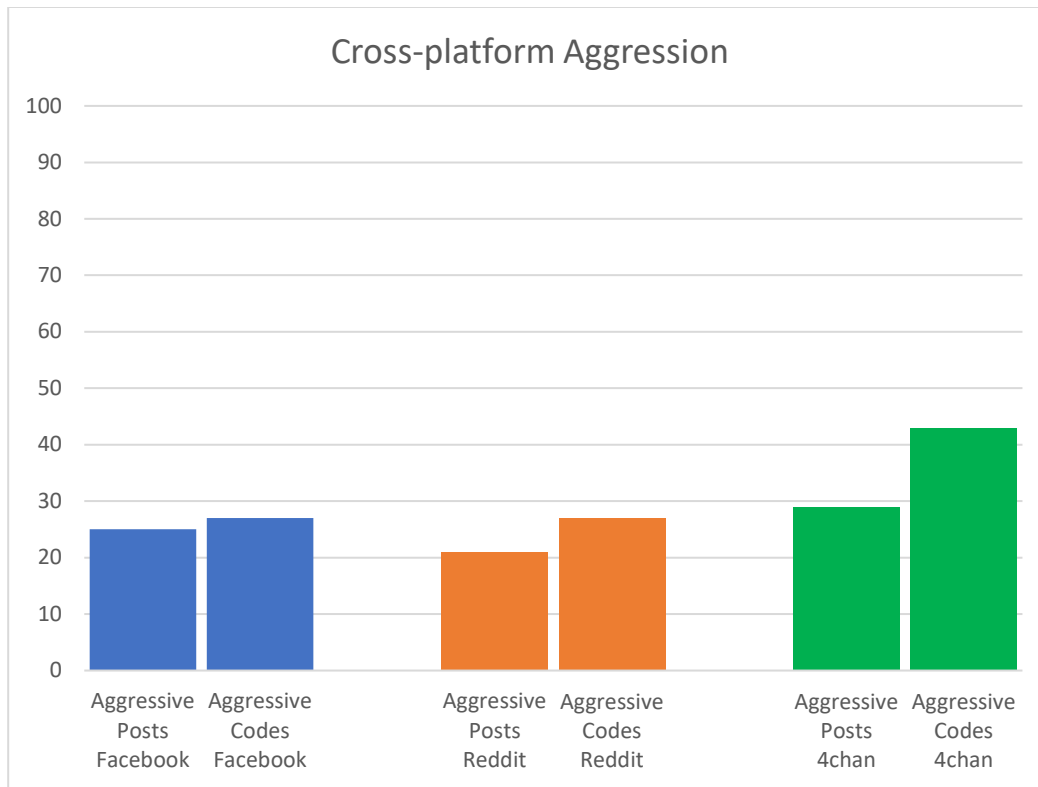
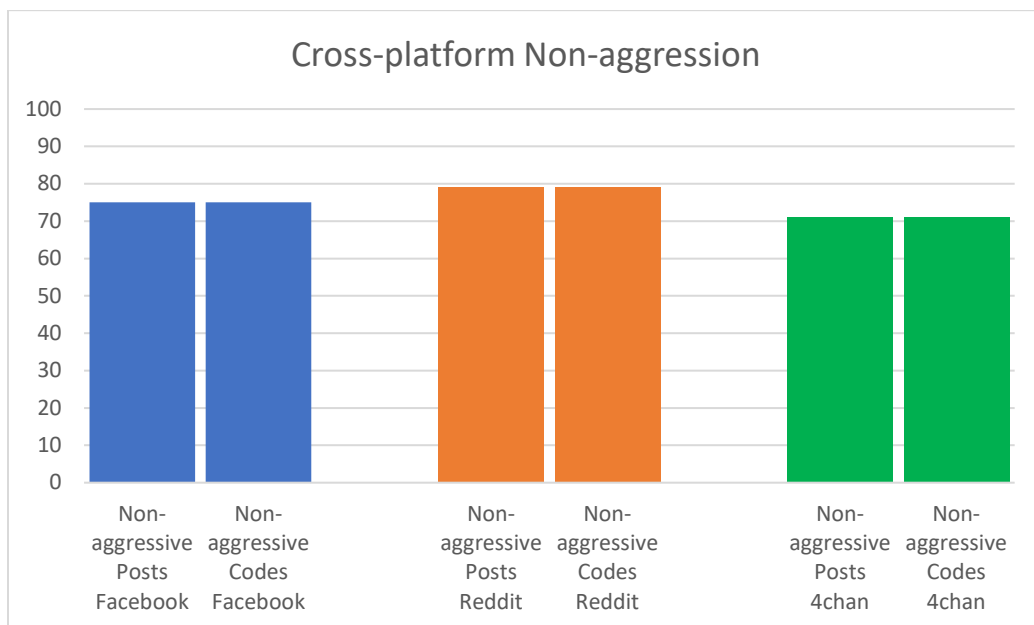Figure 14: Cross-platform Aggression

Figure 15: Cross-platform Non-aggression

Recognizing the challenges inherent in translating qualitative values into quantitative data, these ratios nonetheless provide some insight into user behaviour that would not be seen if the number aggressive posts were analyzed independently of the number of duplicate codes. The data show that the heightened anonymity of 4chan makes users more likely to post aggressive content, as well as more likely to be more aggressive in each comment than on Facebook or Reddit. This finding reinforces work done by previous researchers in the field of anonymity and behaviour, such as Zimbardo (1969), Suler (2004), Wright (2013), and many others. For example, Zimbardo (1969) found that people would present higher levels of aggression in situations where they were unidentifiable and unwatched. Suler (2004) theorized that online disinhibition, caused by anonymity, would have both prosocial and antisocial effects on users' actions. Wright (2013) found that anonymity related positively to cyber-aggression in young adults, and that participants were much more likely to post harmful, antisocial content online when anonymous. My own findings reinforce these three examples, as well as other sources captured within my literature review.

Taxonomic Coding: Diving Deeper into Aggression and Non-Aggression

Saldaña's (2013) qualititative coding framework suggests a taxonomic approach to coding, which includes breaking down an overarching domain into distinct sub-categories to provide a wider breadth of analysis. In addition to studying the effects of anonymity on users' likelihood to post aggressive content online and the level of aggression shown per post, my research also looks at how anonymity affects the type of aggression or non-aggression displayed by users. This relationship is explored in the next two subsections. This deeper look into each platform's aggression and non-aggression domains shows that individual user behaviour is quite

varied. Each platform has unique patterns of behaviour that could be influenced by platform policies, moderation, and/or technical structures.

Aggression Analysis

Earlier in this chapter, I discussed the total volume of aggressive and non-aggressive posts and codes within the data. Within the next two sections, I diver deeper into the taxonomies of each domain (aggresion/non-aggression) to develop a rich data analysis. First, I created two categories within the "aggression" domain based on what the content of the post was directed at. Comments can either be directed toward a user — *"you do realise that we don't really have a gun culture in the UK? What the fuck do you need assault rifles for??!"* (Facebook) — or a more generic direction — *"We have a tyrannical government happening right now. Precious little armed rebellion going on. I call bullshit."* (Facebook). The decision to separate these data into two distinct groupings (user-directed or general/topic-related) was done to add a layer of depth to the analysis by creating a profile of user intention on each platform.

My data analysis indicated that Facebook users display much higher levels of user-directed aggression than users of Reddit or 4chan. This may suggest a possible association in this case between anonymity and user-directed aggression. Out of 27 total aggressive codes on Facebook, 15 (55.6%) were directed at a generic object, while 12 (44.4%) at another user. Out of 27 total aggressive codes on Reddit, 23 (85.2%) were directed at a generic object and 4 (14.8%) at another user. And out of 42 total aggressive codes on 4chan, 39 (92.9%) were directed at a generic object and 3 (7.1%) at another user. While higher degrees of anonymity granted to users led to a higher overall aggression count, both in terms of posts and codes, it appeared to potentially deter them from attacking other users.

One hypothesis is that the structure of Facebook may influence heightened levels of user-directed aggression. On Facebook, users can easily reply to a another user or tag a specific user in their comment, creating more direct person-to-person dialogue when compared to the other platforms analyzed. However, Reddit has similar features (e.g. the ability to reply to specific users), but the same relatively high levels of aggression were not seen there. Another potential influence that ties to the structure of Facebook is that it requires users to attach their real names to their comments. Because of this, Facebook users may feel more personal attachment to what they post, which may be why they react negatively towards users who may hold different opinions or disagree with the content they post.

Within each of the two sub-domains (user-directed aggression and general/topic-directed aggression), I created five taxonomies, based inductively on existing literature, as well as deductive after looking at my data (Suler, 2004; Cho & Kim, 2012; Omernick & Sood, 2013; Gibson, 2019). The five types of aggression coded for were gender/sexual slurs, racism, swearing, threats/violence, and a general aggression category that captured other forms of aggressive behaviour. Each of these was applied to both sub-domains, creating 10 unique codes that are displayed in Figures 16-18.
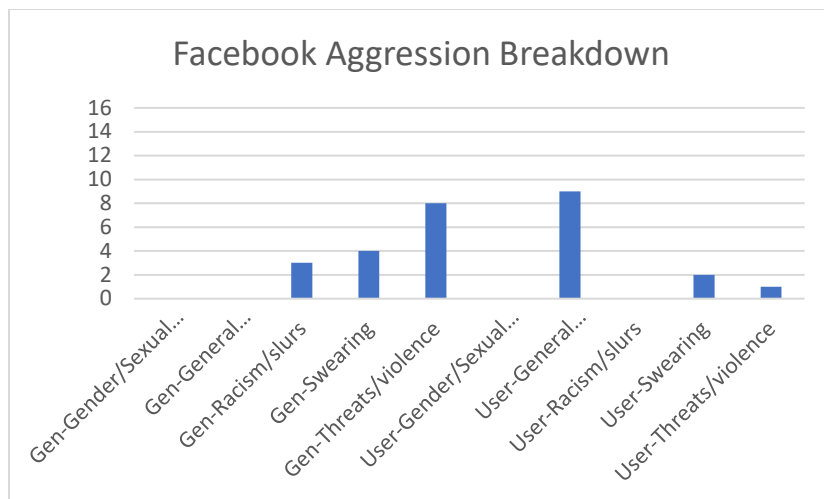
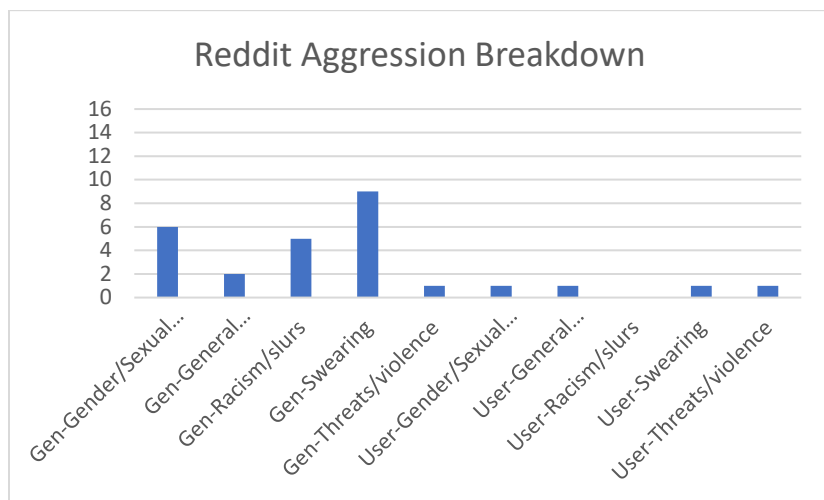Figure 16: Facebook Aggression Breakdown

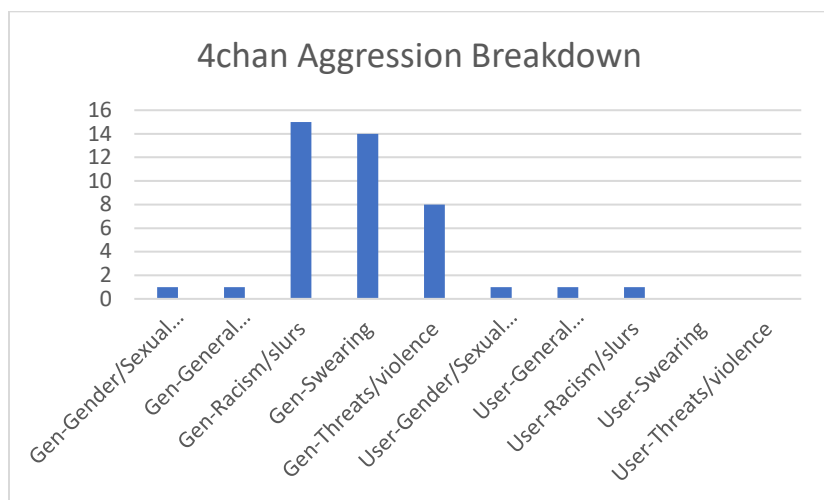Figure 17: Reddit Aggression Breakdown

Figure 18: 4chan Aggression Breakdown

As illustrated in the Figures above, each of the three platforms produced significantly different results of the types of aggression displayed by users. As previously mentioned, Facebook shows the highest level of user-directed aggression, and it also is tied with 4chan for the highest level of generic threats or violence. Reddit shows the highest levels of sexual/gender-based aggression, and 4chan shows the highest level of both racism and swearing. Figures 19-21 below show examples of these prominent types of behaviour on each platform.
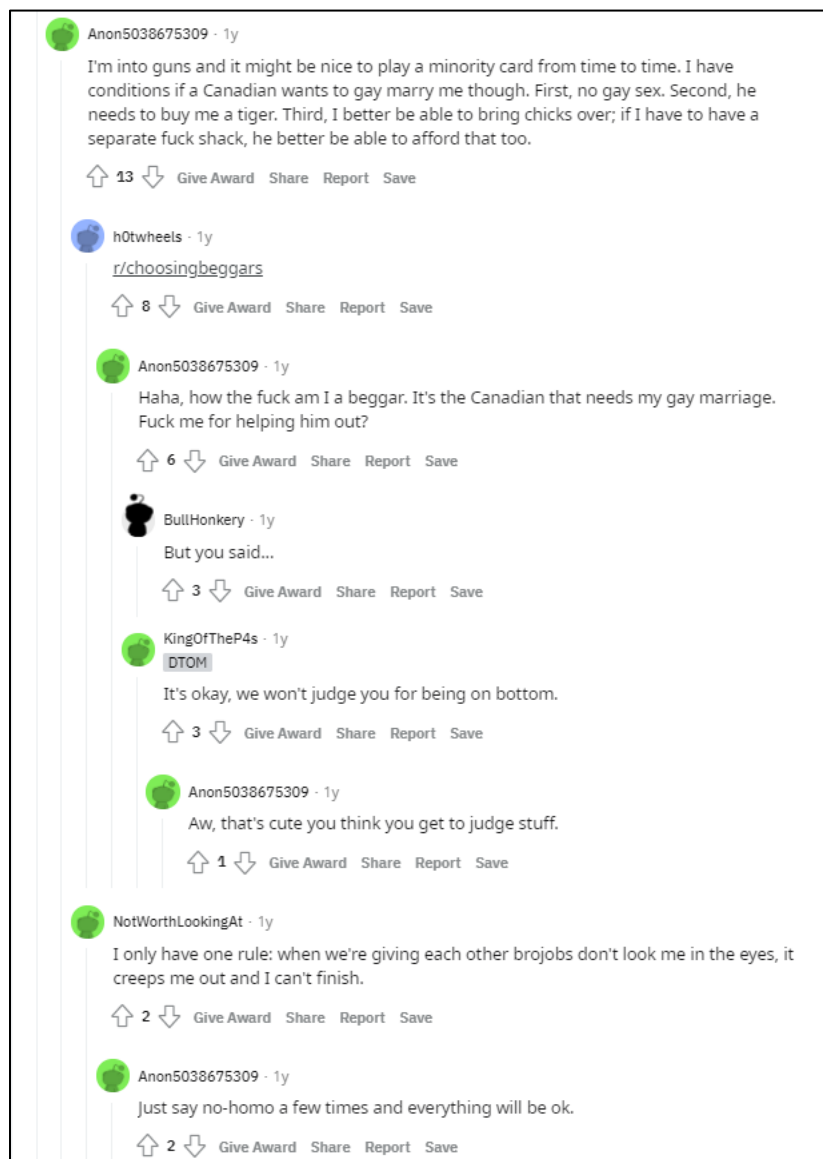

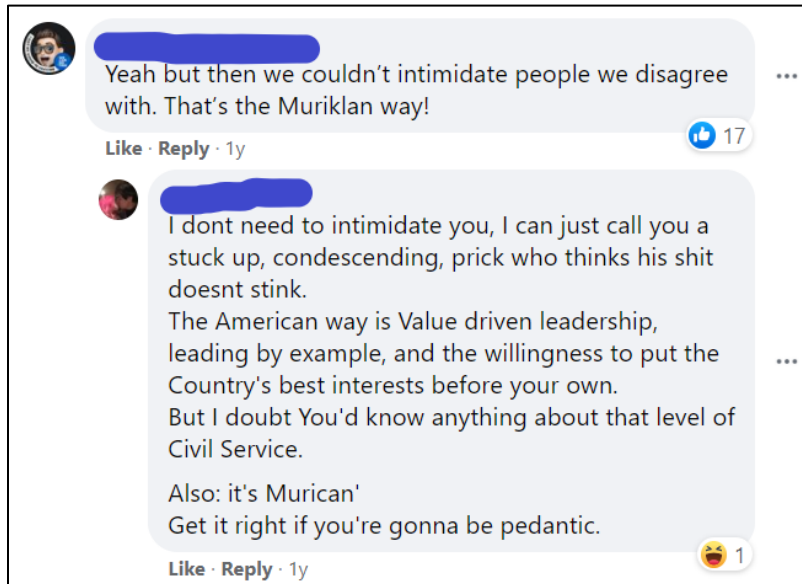Figure 19: Sexual/gender-based Aggression on Reddit

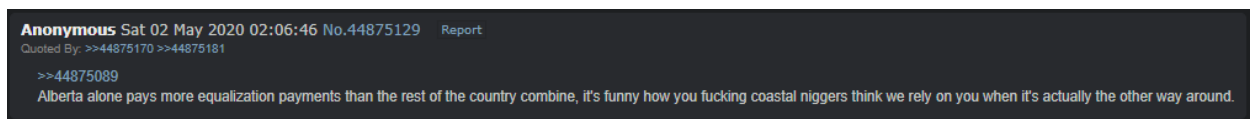Figure 20: User-directed Aggression on Facebook

Figure 21: Swearing and Racism-based Aggression on 4chan

The theory of Social Shaping of Technology would agree that these discrepancies in aggression types can be explained by each platform's structure. Of particular relevance would be the policies on what kind of content is allowed on the platform, and in particular, how each platform enforces these policies through technical and social ways. Facebook's policy on hate speech clearly articulates what types of content the platform does not allow. The platform's moderators, both automatic and human, remove content that clearly violates this policy. This helps explain why there are lower levels of overall aggressive content. However, the policy clearly states that hate speech is against people, rather than concepts or institutions. This might indicate that a content analysis would show that there are fewer examples of user-directed aggression than generic aggression. However, this was not the case in this research project.

"We define hate speech as a direct attack against people — rather than concepts or institutions— on the basis of what we call protected characteristics: race, ethnicity, national origin, disability, religious affiliation, caste, sexual orientation, sex, gender identity and serious disease. We define attacks as violent or dehumanizing speech, harmful stereotypes, statements of inferiority, expressions of contempt, disgust or dismissal, cursing and calls for exclusion or segregation. We also prohibit the use of harmful stereotypes, which we define as dehumanizing comparisons that have historically been used to attack, intimidate, or exclude specific groups, and that are often linked with offline violence" (Facebook Community Standards).

The first rule of Reddit's content policy states that any aggressive content based on identity or vulnerability will not be tolerated. This may explain why we see relatively low levels of user-directed aggression. However, it raises a question around why there are high levels of generic aggression, especially gender- or sexuality-based aggression. On Reddit, users create a persona via a pseudonymous username that is traceable but not personally identifiable. Past behaviour is associated with this account, but for many users, this has no implication on their offline lives. Therefore, there is a level of security granted that may embolden users to partake in more socially unacceptable aggression, such as homophobia and sexism.

"Remember the human. Reddit is a place for creating community and belonging, not for attacking marginalized or vulnerable groups of people. Everyone has a right to use Reddit free of harassment, bullying, and threats of violence. Communities and users that incite violence or that promote hate based on identity or vulnerability will be banned" (Reddit Content Policy).

In stark opposition to the more directive policies of Facebook and Reddit, 4chan's only rule on content is: "You will not upload, post, discuss, request, or link to anything that violates local or United States law" (4chan Rules). The content analysis of this research project confirms that there appears to be little to no removal of content on 4chan based on aggressive behaviour. This lack of moderation and the anonymity provided by the platform increase the abilities of its users to post content that is much more aggressive than is seen on platforms with less anonymity and more content moderation. For example, Racism is especially rampant on 4chan, accounting

for 15 out of 29 aggressive posts (51.7%). Perhaps this behaviour is enabled by the platform's structure. Or perhaps a void existed in digital platforms for users to express this type of behaviour, and 4chan fills the gap. The Social Shaping of Technology perspective would say both potential explanations are true; user behaviour influences technological developments, and technological developments influence user behaviour. Additionally, it is interesting to note that that 4chan has the lowest number of aggressive posts directed at another user of all three platforms. While there were high levels of racism, threats, and swearing on 4chan, it was hardly ever directed towards another user. This finding would be excellent subject matter for future research.

Non-aggression Analysis

In addition to studying the aggression displayed by users, I wanted to code different types of non-aggression to identify any patterns across or within the platforms. The data I analyzed showed that the vast majority of non-aggressive comments across all three platforms are classified as benign comments, with the asking and answering of questions following behind that. The platform with the highest level of anonymity, 4chan, produced the greatest variety of non-aggressive comments and is the only platform that produced any conmments with three of the codes (defending user, general compliment, and user compliment). Figures 22-24 below show the results of the domain/taxonomy coding.
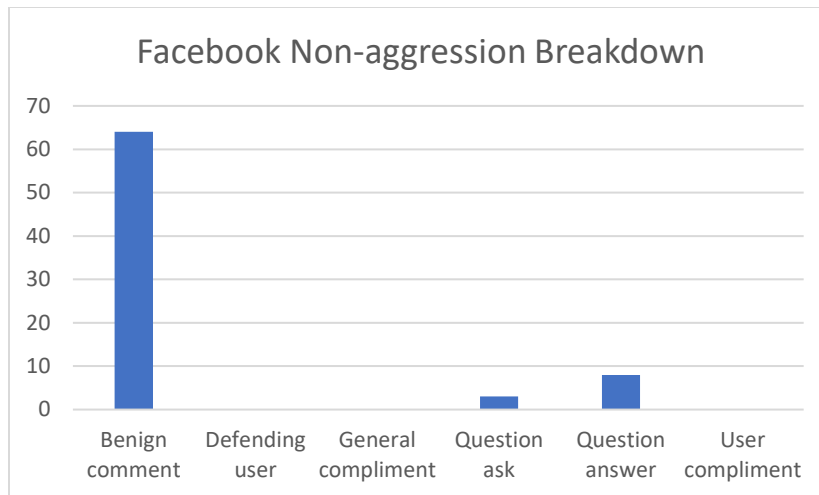
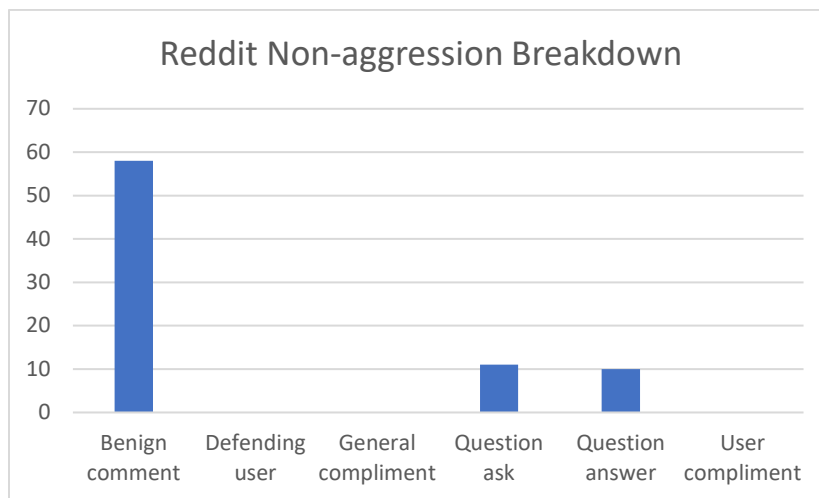Figure 22: Facebook Non-aggression Breakdown
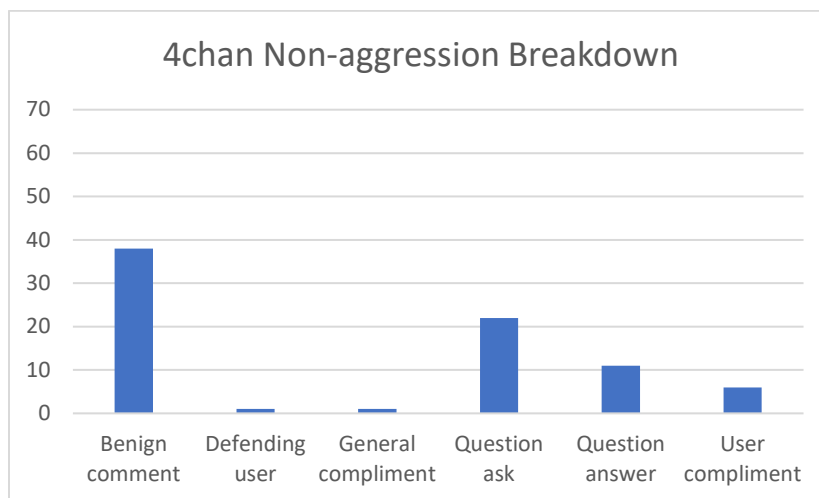

Figure 23: Reddit Non-aggression Breakdown

Figure 24: 4chan Non-aggression Breakdown

This non-aggression analysis did not provide as rich of data as the aggression analysis, but it is still important to review. All the platforms produced a similar number of non-aggressive posts: 75 on Facebook, 79 on Reddit, and 79 on 4chan. Another similarity is the most common code on each platform was what I called a 'benign comment,' which was any non-aggressive comment that did not fit into one of the other five categories. Outisde of the benign comment category, both Facebook and Reddit only produced content that was coded as either a 'question ask' or 'question answer.' On both platforms, the frequency of these two codes was much lower than on 4chan, the most anonymous platform. 4chan also was the only platform on which I saw instances of compliments, directed at either another user or a general topic, or someone defending another user.

The findings from my analysis of non-aggressive user behaviour reinforces claims that online anonymity can induce pro-social behaviours (Lapidot-Lefler and Barak, 2015). In their research, Lapidot-Lefler and Barak performed a similar style of qualitative analysis as I have, assigning phrase-based codes to data, such as "intent to help the other … complimenting the other … and expressions indicating a positive social or amicable atmosphere" (2015, p. 7). This finding supports the claim that anonymity can acutally produce more postive social interactions. Suler (2004) lists some benign effects of online anonymity, including "an attempt to better understand and develop oneself, to resolve interpersonal and intrapsychic problems or explore new emotional and experiential dimensions to one's identity" (p. 321). It could also present itself when people "share personal things about themselves and show acts of kindness and generosity" (Santana, 2014, p. 22). Non-aggressive, prosocial behaviour was not the initial focus of my research project, but after looking at the exisitng literature, it became clear that anonymity influences both antisocial and prosocial behaviours, so focusing on only one or the other would

limit my research. As it turns out, there were some interesting findings in this section, namely that the more anonymous a user is, the more likely they are to present prosocial, non-aggressive behaviour.

Summary

After completing the research and data analysis, it was clear that while anonymity may not have a large impact on the likelihood to post aggressive posts online, it does impact the type and severity of aggression that users display. Each platform's structure and policies also influenced the ways users behaved. Facebook, the least anonymous platform, ranked second for the number of aggressive posts, but had the lowest ratio of aggressive codes to posts, making it the platform with the least severe aggression. However, it did outperform the other two platforms in user-directed aggression, generating an interesting finding that decreased anonymity may correspond to increased user-on-user aggression. Reddit, a platform that provides a pseudonymous structure for users and a mid-range level of anonymity, performed stably across the board, generating mid-level results in almost all comparison fields. It was, however, responsible for the most gender/sexual-based aggression, which may be a result of the platform's culture, not captured within this project. Perhaps the most intriguing findings of this research project surround the most anonymous platform, 4chan. It produced the highest number of aggression codes, the most severe aggression, the least user-directed aggression, the highest number of non-aggression posts (tied), and the most varied non-aggression codes. These findings support much existing literature that states that anonymity directly influences user behaviour online, both prosocially and antisocially. The findings of my research directly answer the first portion of my research question: "How do anonymity, policy, and structure affect social media users' behaviour online, and how does user behaviour in turn shape platform structure?"

However, I can only speculate on the second portion regarding how user behaviour shapes platforms. One example that seems likely to be true is the way 4chan users enter the post ID of the comment they wish to reply to, in the absence of the nested reply system like Reddit and Facebook possess.

      The next and final chapter of my Capstone project summarizes the entirety of this project, lists limitations and declarations, and provides suggestions for further research.

Chapter 5: Project Conclusion

Since the beginning of this MACT program, I have been interested in topics of internet governance and online anonymity. As I progressed through the courses, I focused my potential research areas to either platform moderation or anonymous user behaviour. I believed that anonymity online had the potential to create more harm than benefit, so I set out to research this concept. Through the literature review portion of this Capstone, I discovered that many researchers were also interested in these areas. The relationship between anonymity and social effects, specifically harmful, or toxic, effects, has been studied as far back as the 1960s, and as recently as 2020. There are also examples of studies that look at the prosocial, or benign, effects of anonymity, particularly in the digital sphere. Within the sphere of internet governance, many sources were reviewed that looked at how platform moderation, content policies, and technical structures influenced user behaviour, and I added to this knowledge base within my project. I also wanted to research steps the Canadian government was or was not taking to protect users online and found very little within my literature review. Canada has taken steps to reduce crime and protect minors online, but has not taken steps towards reducing online aggression, which leads to cyberbullying and hate. This inaction is likely intentional, however, as any steps towards censorship online could encroach on users' rights to free speech. In the time since this research project was completed, the Canadian government has taken steps to pass Bill C-10, which would grant the federal government more control over internet governance (Raman-Wilms & Curry, 2021). The Bill is marketed as a tool to help promote and support Canadian content online, but it could potentially give the government too much power over what content is shown to Canadians.

To explore these issues in my research, I performed a qualitative content analysis on 100 comments drawn from each of three popular digital platforms with varying structures, policies,

and levels of user anonymity: Facebook, Reddit, and 4chan. My research found that anonymity did not always directly cause higher or lower levels of aggression. However, but I did discover a few interesting correlations, highlighted in Table 7 below. These findings support much of the existing literature that states that anonymity directly influences user behaviour online, both prosocially and antisocially.

	Anonymity level	Aggressive posts	Aggressive codes	Non-aggressive posts/codes	Most common aggression
Facebook	Least	Second	Second	Second	User-directed
Reddit	Mid	Third	Second	First	Gender/Sexual
4chan	Most	First	First	First	Racism/Slurs

Table 7: Research Findings Highlights

Limitations

Several variables have, or could have, affected the outcome of this research. This study may have benefited from a larger sample size to better understand user behaviour and increase the validity of data analyzed. Many of the 100 comments from each platform were unrelated to the study's case study topic. Similarly, data could have been compared across multiple case studies, rather than just one. Within the coding process, the number of codes I used may have limited the range of data captured and therefore analyzed. Additionally, qualitative coding is influenced by the researcher's own biases, so having multiple coders could have increased the reliability and validity of this study. Facebook was used as an example of a non-anonymous platform, but it has been documented that there are many anonymous and bot profiles (Pourghomi et al., 2020). This limitation may have influenced the analysis, as Facebook may not have been the best source for non-anonymous data. Finally, much of my research borders on the debate of free speech versus censorship, a major contextual element that was not covered within this project.

Declarations

I declare no received funding or conflicts of interest related to this research.

Further Research

After analyzing the data, I identified several items within this project that could benefit from further research. From the literature reviewed, I learned about the South Korean Real Name Verification Law (Cho et al., 2012; Cho and Kwon, 2015); I wonder if a similar law would be accepted in Canada. Now that Bill C-10 is moving forward, this would be an interesting research topic. Within my findings, three subjects stand out as needing more research:

1. Why did users on 4chan, the most anonymous platform, display more frequent and more severe aggression, but also high levels of non-aggression, with the least amount of user-directed aggression?

2. Why did Reddit have more instances of gender/sexual-based aggression, with similar platform policies as Facebook, and a moderate level of anonymity?

3. Why did Facebook have such a significantly higher level of user-directed aggression?

Final Thoughts

This study contributes to online anonymity and privacy literature and can shed light on useful implications to policy makers by examining people's online behaviour on different social media platforms to explore whether content posted on platforms with higher levels of user anonymity users differs from content posted on platforms with lower levels of user anonymity. The findings of my research are also useful to the average social media user. For me, knowing that Facebook users are more likely to demonstrate user-directed aggression will make me pause

and think before engaging on the platform. And knowing that 4chan users are both the most prosocial and antisocial provides some cautionary context about what to expect on that platform.

Overall, I appreciated the opportunity to research a topic I found interesting, growing my own knowledge while contributing to the existing literature around online anonymity, user behaviour, and internet governance.

References

4chan Rules. https://4chan.org/rules

Abramson, B. D., & Raboy, M. (1999). Policy globalization and the "information society": A

view from Canada. *Telecommunications Policy, 23*(10-11), 775-791. doi:10.1016/s0308-

5961(99)00058-0

Ascher, D. L. & Noble, S. U. (2019). Unmasking hate on Twitter: Disrupting anonymity by

tracking trolls. *Free Speech in the Digital Age*, 170-187. https://doi-

org.login.ezproxy.library.ualberta.ca/10.1093/oso/9780190883591.003.0011

Allan, R. (2017). Hard Questions: Who Should Decide What Is Hate Speech in an Online Global

Community? https://about.fb.com/news/2017/06/hard-questions-hate-speech/

Baym, N. (2015). *Personal Connections in the Digital Age.* Cambridge: Polity Press.

BBC. (2017). Amanda Todd case: Accused Dutch man jailed for cyberbullying. BBC News.

https://www.bbc.com/news/world-us-canada-39295474

Beck, L. (2013). Anonymous Outs the Man Who Allegedly Drove Amanda Todd to Suicide.

Jezebel. https://jezebel.com/anonymous-outs-the-man-who-allegedly-drove-amanda-

todd-5952080

Bill C-11. First Reading - Digital Charter Implementation Act, 2020 - Parliament of Canada.

https://parl.ca/DocumentViewer/en/43-2/bill/C-11/first-reading

Bill C-13 [Historical]. Protecting Canadians from Online Crime Act.

https://openparliament.ca/bills/41-2/C-13/

Bill C-30 [Historical]. Protecting Children from Internet Predators Act.

https://openparliament.ca/bills/41-1/C-30/

Bernstein, M. S., Monroy-Hernandez, A., Harry, D., Andre, P., Panovich, K., & Vargas, G. G. (2011). 4chan and /b: An analysis of anonymity and ephemerality in a large online community. *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, 50-57.

Canada's Digital Charter in Action: A Plan by Canadians, for Canadians. [Canada's Digital Charter in Action: A Plan by Canadians, for Canadians - Innovation for a better Canada (ic.gc.ca)](ic.gc.ca)

Canadian Centre for Child Protection. Online Safety. [Resources & Research: Online Safety – protectchildren.ca](protectchildren.ca)

Canadian Centre for Cyber Security. https://cyber.gc.ca/en/

Canadian Charter of Rights and Freedoms. Section 2(b). https://www.justice.gc.ca/eng/csj-sjc/rfc-dlc/ccrf-ccdl/check/art2b.html

Ceferin, R., & Meznar, S. (2014). Online hate-speech and anonymous internet comments: How to fight the legal battle in Slovenia? *Annales: Series Historia et Sociologia*, *24*(3), 477–488.

Cho, D., Kim, S., & Acquisti, A. (2012). Empirical analysis of online anonymity and user behaviors: The impact of real name policy. *45th Hawaii International Conference on System Sciences*. https://doi:10.1109/hicss.2012.241

Cho, D., & Kwon, K. H. (2015). The impacts of identity verification and disclosure of social cues on flaming in online user comments. *Computers in Human Behavior*, *51*, 363-372. https://doi:10.1016/j.chb.2015.04.046

Chang, J. (2008). The role of anonymity in deindividuated behavior: A comparison of

   deindividuation theory and the social identity model of deindividuation effects (SIDE).

   The Pulse, 6(1), 1-8.

Clark, M. (2021). Reddit's lead GameStop hypebeast is being sued for his role in the stock surge.

   The Verge. https://www.theverge.com/2021/2/17/22287612/reddit-wallstreetbets-

   gamestop-lawsuit-roaring-kitty-gill-market-manipulation

Clark-Gordon, C. V., Bowman, N. D., Goodboy, A. K., & Wright, A. (2019). Anonymity and

   Online Self-Disclosure: A Meta-Analysis. *Communication Reports*, *32*(2), 98-111.

   https://doi:10.1080/08934215.2019.1607516

CTV. (2012, October 15). In wake of Amanda Todd suicide, MPs to debate anti-bullying motion.

   https://www.ctvnews.ca/canada/in-wake-of-amanda-todd-suicide-mps-to-debate-anti-

   bullying-motion-1.995254

CTV BC. (2012). Alleged Amanda Todd tormenter tracked down. British Columbia.

   https://bc.ctvnews.ca/alleged-amanda-todd-tormenter-tracked-down-1.997244

Dean, M. (2012). The Story of Amanda Todd. https://www.newyorker.com/culture/culture-

   desk/the-story-of-amanda-todd

DeNardis, L., Cogburn, D. L., Levinson, N. S., & Musiani, F. (2020). Internet governance as an

   object of research inquiry. *Researching internet governance: Methods, frameworks,*

   futures. Cambridge, MA: The MIT Press.

   https://doi.org/10.7551/mitpress/12400.001.0001

DeNardis, L. & Hackl, A. M. (2015). Internet governance by social media platforms.

   Telecommunications Policy, *39*(9), 761–770. https://doi-

   org.login.ezproxy.library.ualberta.ca/10.1016/j.telpol.2015.04.003

Denscombe, M. (2017). The good research guide: For small-scale social research projects (4th

ed.). London: Open University Press.

Department of Justice. Modernizing Canada's Privacy Act. https://www.justice.gc.ca/eng/csj-

sjc/pa-lprp/modern.html#s3

Diener, E. (1979). Deindividuation, self-awareness, and disinhibition. *Journal of Personality and

Social Psychology*, *37*(7), 1160-1171. https://doi:10.1037/0022-3514.37.7.1160

Facebook Community Standards.

https://www.facebook.com/communitystandards/objectionable_content

Facebook Terms of Service. https://www.facebook.com/terms.php

Fact Sheet: Digital Charter Implementation Act, 2020. Innovation for a better Canada.

https://www.ic.gc.ca/eic/site/062.nsf/eng/00119.html

Fraser, N. (1990). Rethinking the Public Sphere: A Contribution to the Critique of Actually

Existing Democracy. *Social Text*, *25/26*, 56-80. https://doi-

org.login.ezproxy.library.ualberta.ca/10.2307/466240

Gee, J. (2005). Semiotic social spaces and affinity spaces: From The Age of Mythology to

today's schools. In D. Barton & K. Tusting (Eds.), *Beyond Communities of Practice:

Language Power and Social Context (Learning in Doing: Social, Cognitive and

Computational Perspectives)*. Cambridge: Cambridge University Press, 214-232.

doi:10.1017/CBO9780511610554.012

Geist, M. (2016). The policy battle over information and digital policy regulation: A Canadian

perspective. *Theoretical Inquiries in Law, 17*(2). doi:10.1515/til-2016-0016

Gibson, A. (2019). Free Speech and Safe Spaces: How Moderation Policies Shape Online

Discussion Spaces. *Social Media + Society*, *5*, 1-15. https://doi-

org.login.ezproxy.library.ualberta.ca/10.1177/2056305119832588

Gillespie, T. (2018). Custodians of the internet: Platforms, content moderation, and the hidden

decisions that shape social media. Yale University Press.

Grimmelmann, J. (2015). The Virtues of Moderation. *Yale Journal of Law and Technology*,

17(1), 42-109. https://doi:10.31228/osf.io/qwxf5

Habermas, J., Lennox, S., & Lennox, F. (1974). The Public Sphere: An Encyclopedia Article

(1964). *New German Critique, 3*, 49-55. https://doi:10.2307/487737

Hackett, R. A., & Anderson, S. (2011). Democratizing Communication Policy in Canada: A

Social Movement Perspective. *Canadian Journal of Communication, 36*(1).

doi:10.22230/cjc.2011v36n1a2459

Horsman, G. (2016). The challenges surrounding the regulation of anonymous communication

provision in the United Kingdom. *Computers & Security*, *56*, 151–162. https://doi-

org.login.ezproxy.library.ualberta.ca/10.1016/j.cose.2015.06.005

Jane, E. A. (2015). Flaming? What flaming? The pitfalls and potentials of researching online

hostility. *Ethics and Information Technology, 17*(1), 65-87. https://doi:10.1007/s10676-

015-9362-0

Jordan, T. (2019). Does online anonymity undermine the sense of personal responsibility?

Media, *Culture & Society*, *41*(4), 572-577. https://doi:10.1177/0163443719842073

Jurich, C. (2019). Mediating dialogue in public online affinity spaces. *Negotiating place and

space through digital literacies: Research and practice,* 21-40*. Information Age

Publishing.

Kasakowskij, R., Friedrich, N., Fierkiewicz, K., & Stock, W. (2018). Anonymous and non-anonymous user behavior on social media: A case study of Jodel and Instagram. *Journal of Information Science Theory and Practice*, *6*, 25-36. https://doi:10.1633/JISTaP.2018.6.3.3

Kim, K. K., Lee, A. R., & Lee, U. K. (2019). Impact of anonymity on roles of personal and group identities in online communities. *Information & Management*, *56*(1), 109-121. https://doi-org.login.ezproxy.library.ualberta.ca/10.1016/j.im.2018.07.005

Kurek, A., Jose, P. E., & Stuart, J. (2019). 'I did it for the LULZ': How the dark personality predicts online disinhibition and aggressive online behavior in adolescence. *Computers in Human Behavior*, *98*, 31–40. https://doi-org.login.ezproxy.library.ualberta.ca/10.1016/j.chb.2019.03.027

Lang, G., Subramanian, R., & Mavlanova, T. (2014). Perceptions and Use of Anonymous Communication across Cultures. *Journal of International Technology & Information Management 24*(1), 53-64.

Lapidot-Lefler, N., & Barak, A. (2012). Effects of anonymity, invisibility, and lack of eye-contact on toxic online disinhibition. *Computers in Human Behavior*, *28*, 434-443.

Lapidot-Lefler, N., & Barak, A. (2015). The benign online disinhibition effect: Could situational factors induce self-disclosure and prosocial behaviors? *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, *9*(2). https://doi:10.5817/cp2015-2-3

Lea, M., Spears, R., & Groot, D. D. (2001). Knowing Me, Knowing You: Anonymity Effects on Social Identity Processes within Groups. *Personality and Social Psychology Bulletin*, *27*(5), 526-537. https://doi:10.1177/0146167201275002

Lowry, P. B., Zhang, J., Wang, C., & Siponen, M. (2016). Why Do Adults Engage in

    Cyberbullying on Social Media? An Integration of Online Disinhibition and

    Deindividuation Effects with the Social Structure and Social Learning Model.

    Information Systems Research, *27*(4), 962-986. https://doi:10.1287/isre.2016.0671

MacKenzie, D. A., & Wajcman, J. (1999). Introductory Essay: The social shaping of technology.

    The social shaping of technology (2nd ed., 3-27). Buckingham, PA: Open University

    Press.

Mahrenbach, L. C., & Mayer, K. (2020). Framing Policy Visions of Big Data in Emerging

    States. *Canadian Journal of Communication*, *45*(1), 129–141.

    https://doi.org/10.22230/cjc.2020v45n1a3471

Meyrowitz, J. (1986). No sense of place: the impact of electronic media on social behavior.

    Retrieved from https://hdl-handle-net.login.ezproxy.library.ualberta.ca/2027/heb.31832.

Miljure, B., & Mangione, K. (2021). Amanda Todd case: Man accused in cyberbullying of B.C.

    teen extradited to face charges. British Columbia. https://bc.ctvnews.ca/amanda-todd-
case-man-accused-in-cyberbullying-of-b-c-teen-extradited-to-face-charges-1.5297727

Milosevic, T. (2016). Social media companies' cyberbullying policies. *International Journal of

    Communication (Online)*, *10*, 5164-5185.

    https://link.gale.com/apps/doc/A478974407/LitRC?u=edmo69826&sid=LitRC&xid=39a
d6d91

Mikal, J. P., Rice, R. E., Kent, R. G., & Uchino, B. N. (2016). 100 million strong: A case study

    of group identification and deindividuation on Imgur.com. *New Media & Society*, *18*(11),

    2485-2506. https://doi:10.1177/1461444815588766

Moderator Guidelines for Healthy Communities. https://www.redditinc.com/policies/moderator-guidelines

Moll, M., & Shade, L. R. (2011). Ten Internet rights and principles. *The Internet tree: the state of telecom policy in Canada 3.0*. Ottawa [Ont.]: Canadian Centre for Policy Alternatives.

Moore, M. J., Nakano, T., Enomoto, A., & Suda, T. (2012). Anonymity and roles associated with aggressive posts in an online forum. *Computers in Human Behavior*, *28*(3), 861-867. https://doi:10.1016/j.chb.2011.12.005

Office of the Privacy Commissioner of Canada. Online Privacy. https://www.priv.gc.ca/en/privacy-topics/technology/online-privacy-tracking-cookies/online-privacy/

Omernick, E., & Sood, S. O. (2013). The Impact of Anonymity in Online Communities. *2013 International Conference on Social Computing*. https://doi:10.1109/socialcom.2013.80

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, *17*(11), 776-783. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=ovfta&NEWS=N&AN=00000487-196211000-00005

Perbawani, P. S., Rahayu, R., & Anshari, I. N. (2018). Netizens Anonymity in Indonesia's Digital Democracy: Political Participation in Social Media According to the Online Disinhibition Effect Theory. *PCD Journal*, *6*(2), 185. https://doi:10.22146/pcd.41905

Pfitzmann, A., & Köhntopp, M. (2001). Anonymity, Unobservability, and Pseudonymity — A Proposal for Terminology. *Designing Privacy Enhancing Technologies Lecture Notes in Computer Science*, 1-9. https://doi:10.1007/3-540-44702-4_1

Phys.org. (2018). S*panish consumer group to sue Facebook over data sharing.* Phys.org.

https://phys.org/news/2018-05-spanish-consumer-group-sue-facebook.html.

Pourghomi, P., Dordevic, M. & Safieddine, F. (2020). Facebook fake profile identification:

technical and ethical considerations. *International Journal of Pervasive Computing and*

Communications, *16*(1), 101-112. https://doi-

org.login.ezproxy.library.ualberta.ca/10.1108/IJPCC-06-2019-0049

Raman-Wilms, M., & Curry, B. (2021). What is Bill C-10 and why are the Liberals planning to

regulate the internet? *The Globe and Mail*.

https://www.theglobeandmail.com/politics/article-what-is-bill-c-10-and-why-are-the-

liberals-planning-to-regulate-the/

Reagle, J. M. (2015). Reading the comments: Likers, haters, and manipulators at the bottom of

the web. Cambridge, MA: MIT Press.

Reddit Content Policy. https://www.redditinc.com/policies/content-policy

Ribeiro, J. (2015). *Facebook sued in US court for blocking page in India.* Computerworld.

https://www.computerworld.com/article/2930874/facebook-sued-in-us-court-for-

blocking-page-in-india.html.

Rösner, L., & Krämer, N. C. (2016). Verbal Venting in the Social Web: Effects of Anonymity

and Group Norms on Aggressive Language Use in Online Comments. *Social Media +*

Society, *2*(3), 205630511666422. https://doi:10.1177/2056305116664220

Royal Canadian Mounted Police. Digital Policing Strategy: The Connected RCMP.

https://www.rcmp-grc.gc.ca/en/connected-rcmp

Saldaña, J. (2013). *The coding manual for qualitative researchers* (2nd ed.). Los Angeles, CA:

SAGE.

Santana, A. D. (2013). Virtuous or Vitriolic. *Journalism Practice*, *8*(1), 18-33.

>   https://doi:10.1080/17512786.2013.813194

Savage, P. (2008). Gaps in Canadian Media Research: CMRC Findings. *Canadian Journal of*

>   *Communication*, *33*(2), 291–301. https://doi.org/10.22230/cjc.2008v33n2a2060

Seigfried-Spellar, K. C., & Lankford, C. M. (2018). Personality and online environment factors

>   differ for posters, trolls, lurkers, and confessors on Yik Yak. *Personality and Individual*

>   *Differences*, *124*, 54-56. https://doi:10.1016/j.paid.2017.11.047

Silke, A. (2003). Deindividuation, anonymity, and violence: Findings from Northern Ireland. *The*

>   *Journal of Social Psychology*, *143*(4), 493-499.

Simona, M., Regan, K., Herrera, F., Favero, L., & Quaresema, H. (2019). Research into online

>   platforms' operating models and management of online harms (Rep.). London: ICF

>   Consulting Services Limited. Retrieved from

>   https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_
>   data/file/811448/Research_into_online_platforms_operating_models_and_management_
>   of_online_harms.pdf

Sparby, E. M. (2017). Digital social media and aggression: Memetic rhetoric in 4chan's

>   collective identity. *Computers and Composition*, *45*, 85–97. https://doi-
>   org.login.ezproxy.library.ualberta.ca/10.1016/j.compcom.2017.06.006

Suh, K.-S., Lee, S., Suh, E.-K., Lee, H., & Lee, J. (2018). Online Comment Moderation Policies

>   for Deliberative Discussion-Seed Comments and Identifiability. *Journal of the*

>   *Association for Information Systems*, *19*(3), 182–208. https://doi-
>   org.login.ezproxy.library.ualberta.ca/10.17705/1jais.00489

Suler, J. (2004). The Online Disinhibition Effect. *CyberPsychology & Behavior*, *7*(3), 321-326.

> https://doi:10.1089/1094931041291295

TheSomebodytoknow. (2012). My story: Struggling, bullying, suicide, self harm. YouTube.

> https://www.youtube.com/watch?v=vOHXGNx-E7E

Wachs, S., Wright, M. F., & Vazsonyi, A. T. (2019). Understanding the overlap between cyberbullying and cyberhate perpetration: Moderating effects of toxic online disinhibition. *Criminal Behaviour and Mental Health*, *29*(3), 179-188.

> https://doi:10.1002/cbm.2116

Williams, K. S. (2006). On-line anonymity, deindividuation and freedom of expression and privacy. *Penn State Law Review*, *3*, 687-701.

Williams, R., & Edge, D. (1996). The social shaping of technology. *Research Policy, 25*, 856-899.

Wright, M. F. (2013). The relationship between young adults' beliefs about anonymity and subsequent cyber aggression. *Cyberpsychology, Behavior, and Social Networking*, 16(12), 858-862. https://10.1089/cyber.2013.0009

Wright, M. F., Harper, B. D., & Wachs, S. (2019). The associations between cyberbullying and callous-unemotional traits among adolescents: The moderating effect of online disinhibition. *Personality and Individual Differences*, *140*, 41-45.

> https://doi.org/10.1016/j.paid.2018.04.001

Wu, S., Lin, T., & Shih, J. (2017). Examining the antecedents of online disinhibition. *Information Technology & People*, *30*(1), 189-209. https://doi:10.1108/itp-07-2015-0167

Young, R., Miles, S. & Alhabash, S. (2018). Attacks by anons: A content analysis of aggressive posts, victim responses, and bystander interventions on a social media site. *Social Media*

+ *Society*, *4*. https://doi-

org.login.ezproxy.library.ualberta.ca/10.1177/2056305118762444

Zhang, K., & Kizilcec, R. (2014). Anonymity in Social Media: Effects of Content

Controversiality and Social Endorsement on Sharing Behavior. *Proceedings of the 8th

International Conference on Weblogs and Social Media*, ICWSM 2014.

Zimbardo, P. G. (1969). The human choice: individuation, reason, and order versus

deindividuation, impulse, and chaos. *Nebraska Symposium on Motivation*, *17*, 237-307.

Zimmerman, A. G., & Ybarra, G. J. (2016). Online aggression: The influences of anonymity and

social modeling. *Psychology of Popular Media Culture*, *5*(2), 181-193.

https://doi:10.1037/ppm0000038

Zuckerberg, M. 2018. A Blueprint for Content Governance and Enforcement.

https://www.facebook.com/notes/mark-zuckerberg/a-blueprint-for-content-governance-

and-enforcement/10156443129621634/?hc_location=ufi