The New Double Standard: How Reddit Moderators struggle under the Reddit Ecosystem

An in-depth case study about Moderators feeling the consequences under the system of the social platform Reddit

Student Name: Maarten van der Stad
Student Number: 474033

Supervisor: René Konig

Media & Business
Erasmus School of History, Culture and Communication
Erasmus University Rotterdam

Master Thesis
June 23rd 2022

The Perspective of different subreddit Reddit moderators on the operations of the Reddit as a website

ABSTRACT

In-depth interviews expose the struggles moderators are dealing with, including bad actors brigading their community in an attempt to get it banned, to the inherent lack of transparency the website displays that affects their mental health. The moderators only have their own ethical and philosophical ideology to refer to when dealing with controversial content and cannot rely on a broken structure. The research question delves into the perspective of these moderators that are stuck in a toxic ecosystem only incentivized to generate profit, leaving informal processes to affect the space unchecked. These informal processes, although some advantages, endanger users on the site and the mainstream, polarizing, radicalizing and potentially pushing political agendas. Law makers and other legal involvements seem to move slowly to have a proper impact on the space and lack the knowledge for a rapid and effective response that requires to be flexible in the rapidly changing online space. Among interviews there also exists the discussion of nuance and internet culture, depicting a discussion around wider problems in society that are imbedded within the content that is posted on some of the subreddits. Interpersonal relationships play a large role depicting power balances in the ecosystem, delegating of the incompetent and apathic to the pleas of change regarding illegal content and user conduct. Naturally, case studies and lived experiences of said case studies become the centre of attention as the main vehicles for constructing a final point. One of the most contentious topic of controversial content online is revealed to be viewed at incorrectly by official institutions which ask the wrong question of where to draw the line and why. One has to apply themselves to the discussion to understand and verify controversiality, requiring empathy, sympathy and knowledge of internet culture. Finally, discussion of bots, auto moderator and the algorithm reveal the undeveloped spaces and networks that are pressured to use those tools for moderation. To accommodate the several themes found within the data, and to make sense of its interpretation, the method thematic analysis is applied by Braun and Clarke (2008). Conclusions drawn from the analysis of data reveal a slow and broken system, ideological fanatical moderators and apathetic corporate shareholders that do not care to see change unless it increases their profit endeavours.

Words: 376

KEYWORDS: *Reddit, Thematic Analysis, Moderation, Controversial, In-depth Interviews*

Table of Contents

Chapter 1

Introduction – 1.0

Social platforms are currently at the forefront for cultural development in the 21st century, increasing in importance annually (Williams, 2019). This is evident by the increase in attention to it in mainstream media and governmental regulation (Williams, 2019). As time progresses and these platforms develop, regulations develop with them (Williams, 2019). However, laws and regulations that are to be implemented suffer from the bureaucratic obstacles that come along the way, and ever more questions arise if the elected individuals with a slow implementation of regulation are set up to understand internet culture sufficiently and if they are equipped to combat content that plague the current affairs of the online public sphere in a rapidly changing environment (De Streel et al., 2020). As several governments, laws and mainstream media discuss what regulations should be implemented on a federal level for SNS platforms, those at the front line moderating and manipulating the content better understand the consequences of the current and potential regulations.

Reddit, the ideal space to observe in tandem with this discussion due to their communal structure, reveals how moderators struggle to incentivize democratic discussion among its members due to the lack of impactful legal actions, the unchecked spread of illegal content and corruption. Within the information age, and younger generations taking to online spaces, several cultures developed (Williams, 2019). In this split to regulate the online, several fractures of understanding have begun to form as some cite it as a lack of nuance due to the new communities that continuously pop-up in SNS spaces.

This prompts a lack of understanding between institutions and actors who oversee regulation and misinterpret content due to lack of nuance. This has already occurred for meme culture, controversial content and internet culture (Weinberg, 2018; Kahne & Bowyer 2018). Many spaces have allowed to become polarized and radicalized due to the failure to realize which users present the actual threat among the community of Reddit, citing several double standards and even a case of illegal content being spread around unchecked. Reddit ideology and moderator culture have gained a reputation as one to be distrusted and those which have erased lines of nuance in the circulation of online content. Ultimately, through semi-structured in-depth interviews, commentary from these very moderators from various communities on Reddit share their experience on the standards and expectations imposed on them by the mentioned actors.

Through this method, the research paper will uncover the different themes that serve as motivators to the decisions moderators make within the given Reddit ecosystem and how it either empowers, or restrains them. This has lead to the research question:

How do Reddit moderators perceive the structure, motivators and actors of the Reddit ecosystem that influence their moderation style and community surroundings?

Chapter 2

Theoretical Framework – 2.0

Before addressing the several actors with motivators, the keyword has to be defined first. Motivators in this context refers to any element or aspect of an actor or structure that can influence the moderating style and/or decisions when moderating content. Despite having a positive connotation, the word applies to both negative and positive aspects, in which for example a moderator can be pushed to moderate with unfavorable elements because their circumstances require to do so. Finding out which motivators and why these influences exist will be done by observing macro and micro elements. The macro include the public sphere, legal boundaries and the Reddit ecosystem. Micro includes subreddits, moderators and the auto-moderator bot. These views full fill part of the entire structure to, theoretically to how moderators would be influenced by their surroundings.

The public sphere – 2.1

The public sphere was first coined by German philosopher and sociologist Jurgen Habermas (1962). According to his theory the function of a 'public sphere' was to bring all people of a society together, for them to discuss societal issues and development to then influence political action to bring upon the required change. In relation to the internet, The public sphere found its theory to come alive online as well (Habermas, 1962, p.12). The internet provided instant connection and convenience, and as tech pioneers and entrepreneurs discovered the space and rapid growing popularity, they quickly started to develop SNS platforms (Schäfer et al., 2018). Although discussions range about the advantages and disadvantages of online public sphere, availability and the convenience of the internet certainly allowed for a wider and more diverse range of demographics to join in the discussion (Schäfer et al., 2018). Habermas was also aware of the phenomenon and further commented on the versatility of the variant, the 'digital public sphere' in which he praised how the values of the public sphere were accentuated in the online environment.

Despite the success it brought upon reviving the theory, praise however also drew heavy criticism; which pointed to ignorance which became prevalent in the digital public sphere. Minority opinions would be ignored and removed and lower classes with lower income would not have the means to afford the technology to join in the discussion. These concerns heavily lean into how the public sphere and its democratic spaces are tied to capitalist institutions and their position to exploit the online audience (Van Dijck et al., 2018; Loader 2017). Loader (2011) and Van Dijck et al. (2018) comment on these economic structures voicing that it

maintains an imperfect democratic online sphere of discussion on purpose for profit. Papacharissi (2002) further builds on critique of the divided classes and expresses worry on the social oppressions that occur in online spaces. Their critique lined up when discussing potential change for the platforms as 3 factors eliminated progress. These are, "The big Five", Core Corporation structures and Datafication.

The big five in the text by Van Dijck et al. (2018) refers to 5 large private companies that dominate the market spaces for information technology, Google, Microsoft, Apple, Facebook and Amazon. These companies dominate the western (North America and Europe) online ecosystem of platforms (Van Dijck et al., 2018, p.12). This corporate monopoly thus forces audiences to use their tools and exploit their data; meaning they have no choice but to submit having their data harvested to access internet services. Thus, as capitalist notions are in place, it causes SNS platforms to focus on a profit driven structure with investors constantly pushing for the monetization of tools on the platform to emit growth (Van Dijck et al., 2018). The current lucrative strategy is to provide services for free in exchange of gathering data. Due to its mass success as shown by the growth margins of tech companies such as Apple, deemed the most valuable company in the world as of 2019, companies have resorted to prioritizing datafication tools such as algorithms and entertainment instead of creating proper platforms of equal discussion.

Combating a flawed structure is nothing new and national governments had already imposed regulations such as the Digital Services Act to limit the reach of private companies data harvesting (Van Dijck et al., 2018). Simultaneously it was done to empower the user as they could have more autonomy of their data (Van Dijck et al., 2018, p.65). In essence, the theory of the public sphere reveals that it exists but has instead become an exploited space for monetization and entertainment, rather than fruitful discussion. Van Dijck et al. (2018) still critiques Habermas's sphere, as a romanticized idea that will never come to fruition. In relation to moderators, its clear they are the actors tied to the flawed space and have to make use of tools that are inadequate to tackle the content posted in a long term basis. Moderators have to operate under the given circumstances and structures bestowed upon them by the administrators of the websites. As long as the objective of profit remains a priority, these spaces are unlikely to undergo large changes and have to instead, understand and work around the inhibitions that come in their existing public sphere.

This was elaborated upon by Loader (2011) as he discussed the power balance between social media and users. Social media is the main medium to connect and communicate with other people, yet is seen as entertainment because is more profitable this way (Fuchs, 2013;

Sevignani, 2015). To reinforce the previously stated notion, social media rather promotes personalized ads and communities for the sole purpose of gathering data that can be sold to advertisers, diluting the opportunity for serious discussion for socio-political change (Loader, 2011, 2017). This does imply societal issues are not discussed on these platforms. However it serves only to incentivize 'slacktivism' as coined by Clark in 1995 and allows corporations to respond accordingly to those discussions by manipulating it to their advantage for their profit structure. The most prominent example includes public relations (PR) (Appendix A).

Slacktivism refers to the participation of a political or social cause with very little commitment or effort, therefore not actually bringing along meaningful change (Clark, 1995). This is also described as virtue signalling, in which attempts are made to support a social movement in order to gain social approval (Clark, 1995) This becomes much easier to do as the internet has made information available globally, and therefore increases the abundance of social issues one can choose to support. SNS platforms have thus provided several filters to increase people's engagement with each other, thus more data, without achieving actual change.

Both Poell et al. (2018), Loader (2011, 2017) emphasize how these large corporations operate their platforms to maximize extraction of data and will do whatever to grow and maintain engagement and determine on that basis what is and is not allowed on to be discussed on the platform other than overt illegal content. Further attempts at regulation have only been achieved in recent years, indicating the novelty of this uncharted phenomenon.

Legal Framework - 2.2

One of the most important motivators to moderation are laws and policies. This is because these are unavoidable rules that cannot be overwritten by websites or private companies. This gives the moderator a semblance of a framework to follow, knowing exactly what content should be removed (De streel et al., 2020). Especially in a capitalist environment in which the internet is set up, due to private properties on the net, it is the only hand that can effectively influence corporations to enact change. Moderation in this sense is the most clear and simple process, for example, if content is defined as illegal, it is removed and the user who posted it faces consequences. If these rules are not followed, the website is removed and then intellectual property owners face the consequences (De Streel et al., 2020, p.30). In other instances, the law may not have as clear rules as various genres of content can display a controversial front, instead of holding a directly illegal element, such as nudity of a minor or violent content of in a terrorist state. This borders on issues such as hate speech. The biggest

advantage to legal institutions on online content is that they are linked to access to the market, and as stated in the previous chapter, the incentive to make profit for private companies will incentivize them to follow regulations. The effectiveness can be noted by the example of the case study referring to the threat of Facebook/Meta to leave the EU market being unsuccessful (Forbes, 2022).

Unfortunately, there is only 1 concrete law in place that affects online content. This is the Digital Services Act (DSA). It focuses on a user autonomy approach in which individuals can flag online illegal content to get it removed by official institutions, and trusted flaggers can garner privileges in channels of official institutions that deal with this content (Europa, 2022). This is achieved by the flagger demonstrating "particular expertise and competence" (Europa, 2022).

The remaining, are policies in place which guide moderators to making the correct decisions when confronted with illegal content. The EU is aware of this issue and also mentions it on their own website, "The Commission is concerned that the removal of illegal content online is not effective enough", which implies they are still developing policies and laws to enforce on the online space. Regulation on media is a slow process, and as the internet is a continuous rapidly changing landscape, it can take up to several years to develop a proper impactful law (De Streel et al. 2020). The current policies in place to combat illegal content is a recommendation on measures, adapted from a previous legal form into a non-binding legal form. This is due to most platforms holding a global position that intersect with other international laws in order to operate in their domestic market. The measurements are set up as 5 rules, 'Clearer notice and action procedure', 'More efficient tools and proactive technologies', 'Stronger safeguards to ensure fundamental rights', 'Special attention to small companies' and 'Closer cooperation with authorities'. These rules and its useful applications are further discussed in the CRFD framework chapter below.

When referring to illegal content, the focus is more highly attuned to intellectual property rights, because its limits are easier to define. Although facing its own debate, the basic premise is that when content is used by an individual which is legally owned by another company or owner, and it is not under the fair use law, it has to be removed unless some type of permission has been authorized (De Streel et al., 2020, p.30). Other forms of illegal content that are more clearly defined those which have historically been illegal in the system already, such as the sexualization of minors and terrorism. Translated to the online space, illegal content is that which depicts the sexualization of minors and terrorist content. At last, the content which produces largest-scale discussion is on banning xenophobic and hate speech content, which

mainly refers to the overt use of slurs or calls for violence towards a group of people (De Streel et al., 2020, p.16). However on this topic, lawmakers, academic scholars struggle to create measures on what is considered hate or xenophobic speech (De Streel et al., 2020, p.18).

This issue resides in the deeper conflict on the term controversial content. Jasser et al., (2021) insinuates that online controversial content specifically focuses on "controversiality in social cascades" (p.1) and that it is an inevitability of SNS platforms. The study by Jasser et al., (2021) was conducted on Reddit and explains that controversial content appears when users have mixed opinions about a comment. Notably, the study had a self-described "novel approach" (p.2) to the term controversial content, with a user-centered perspective on controversiality. The novel approach by the authors was an attempt at theorizing and solving the impasse problem. This is because different users have different levels of tolerance towards controversial content. Due to Reddit being a globally accessible platform it widens the margin for an influx of users with different levels tolerance (Jasser et al., 2021).

To best define the term controversial content from the legal perspective, the focus will shift to laws created by the EU. The EU has published legal documents about the subject and evidence of discussion research and discussion in academic peer-revied journals, providing nuance and credibility to their current reasoning and claims on the subject (Jasser et al., 2021; Guimares & Weikum, 2021). Furthermore these rules can wholefully be applied to the website because westerners make up the largest demographic on the website. In essence, the website would reflect the values according to its demographic. Americans, making up 52% of the platform active userbase as of 2020 (Jasser et al., 2021 p.5), followed by several EU countries and Australia. Notable outliers such as India, Brazil and Philippines are within the top 10 of making up the Reddit demographic however. Reviewing the legal perspective from the EU and the USA does not mean it dismisses the perspective on controversial content or topics from other countries entirely, or that are of no importance. However, it will emphasize that controversial content from the given subreddits will be judged under that lens within this paper. This will understandably view this research as an approach from a more western-centric perspective and bias, yet as the subject group of moderators are either from Europe or the USA, it does signify likely more relevance to abide by these guidelines. Following this, controversial content will also be viewed through the work of Menses (2021), which assigns proper legal criteria to identify and moderate it. This provides a solid foundation to what structure moderators work under, and what remaining content they have to moderate in order to keep a civil online democratic space.

EU law – 2.2.1

The academic and EU parliament official research journal on Illegal content online written by de Streel et al., (2020) that includes a section (2) that explains the exact laws that are put in place against illegal content. "reviews and assesses the EU regulatory framework on content moderation and practices" and further explains that this refers to "key online platforms" (p.3). Thus, this statement applies to Reddit as it is a large online media content platform, similar to Facebook, Instagram and Twitter. An important acknowledgement of the journal is the dissimilarity between several member states of the EU. It explains how it takes into account the diverging rules and the national public administration values that differ significantly among countries. A large negative is that this is causes as a slow and rigorous democratic process to enact and pass laws (De Streel et al., 2020, p.5). However, this does indicate that regulations and laws passed which are enforced have reached a democratic consensus on the topic of controversial content, and therefore authenticates and legitimizes the process. Finally, the academic journal outlines the role of the online platform in enforcing these laws, as it is not only the duty of the member state, but also the commercial companies to enforce an ethical, diverse and respectable online space. The official EU law (2021) defines illegal content by outlining it into 4 types, (i) Child sexual abuse material, (ii) racist and xenophobic hate speech (iii) terrorist content and (iv) content infringing intellectual property rights.

In the cases of (i), (iii) and (iv) the concern of censoring content is not necessarily a widespread issue for moderators, and in fact is one of the sounder reasons to their functioning and existence in the digital sphere. Moderators who censor the content often have to review it in order to justify banning it, which was the case for Facebook moderators (Almerkhi et al., 2020, p.8). Notably, reviewing such content can come at a cost as many suffered from mental health and PTSD complications after moderating online content for approximately a year (Almerkhi et al., 2020, p.8). The moderation of content and spaces does subject moderators to the possibility of obscene content of which official EU and US law are only recently starting to tap into. However, it is a relevant point to bring up as it is an emerging and currently unfamiliar subject in this area (Jasser et al., 2021, p.7).

The ethical framework of moderators begins to impact practices with (ii), racist and xenophobic hate speech. The academic journal states the implementation of a Counter-Racism Framework Decision (CRFD) in 2008 to counter hate speech and racism. Furthermore this issue is widely acknowledged for content online as Code of Conduct in (2016) ensured the CRFD implementation for online communities and spaces on all websites. Despite the official structures set in place, the academic journal outlines the challenging role of defining racist and xenophobic hate speech. This is because this is one of the more widely contested subjects among member states of defining what hate speech is (De Streel et al., 2020, p.30). The fragmented rules on criminal procedures by each member state and their own practices thus make it hard to enforce the framework effectively.

In light of this, there are procedures and rules of moderation enforced in the framework that have to be applied by the main online platforms themselves (De Streel et al., 2020, p.30). Coined as the 5 media commitments, the platforms have to first, draw attention to the type of content that is not allowed on the website. Platforms have to make the active step of promoting community standards and guidelines of which they specifically prohibit the incitement of violence and hateful behavior. Second, the platform must have an official and effective review process in place when dealing with reports and notifications of illegal hate speech, and therefore remove them and make them inaccessible; applying this according to community guidelines/standards and national transposition laws to the majority of the reported illegal hate speech within 24 hours. Third, the staff have to be trained in relation to contemporary societal developments. Fourth, the commercial companies need to encourage the reporting of hate speech, as so done by the role of experts, preferably, in collaboration with Civil Society Organizations (CSOs); therefore, indirectly supporting them. Fifth and final commitment is the strengthening of communication and cooperation between them and national authorities; with a focus on procedures for submitting notifications, and collaborating with other online platforms to improve and ensure the exchange of best practices between them.

Advantages and Disadvantages of CRFD - 2.2.2

The advantages of implementing the CRFD through Code of Conduct has been relatively successful. The commission which receives reports from the, platform's states that since 2019, 88.9% of notifications were reviewed within 24 hours, thus an exponential increase of 40% since 2016, and the speed at which such content was reviewed improved the average to 71.7% of reported illegal hate speech removed (De Streel et al., 2020, p.30).

The disadvantages listed are larger in part connected to the ethical concerns and flaws the statistics or implemented practices may represent. The advantages listed, are a result from interpreting them from face-value (De Streel et al., 2020, p.30). This does not directly invalidate the statistics but does put into question legitimacy and authenticity. As it is stated "However there is little information on how the statistics are calculated" (p.30). Next, platforms still remain ambiguous towards users in regards to notifications and content of illegal hate speech, as only 65.4% of users are receive feedback. The percentage of feedback was higher towards 'trusted flaggers' whom presumably are trusted by the platforms, i.e. moderators and verified users. Finally, the commitments in place incentivized a system in which the focus shifted to increasing the speed and number of removals instead of reviewing the actual illegality of the content (De Streel et al., 2020, p.30; Quintel & Ulrich, 2019).

The academic journal investigated and reported 5 major weaknesses due to the standards of these commitments. The first includes the risk of private censorship practices, by following and prioritizing the application of community guidelines and standards. The second is the lack of precision in determining the validity of a notification. Third, the absence of appeal mechanisms for users whose content has been withdrawn. Fourth, content is not required to be sent and reviewed by competent national authorities when removed according to guideline rules and standards. Last and fifth, the 24-hour window of reviewing comments has been speculated as an impossible goal, due to internet traffic, influx of comments versus number of moderators and the remaining commitments they have to follow; therefore, leading them to over-blocking practices.

This section within the academic journal is notable as it is the only category which includes the speculation of disadvantages and mentions the divergence of rules for the definition of hate speech among member states. It too recognizes the novelty of moderation on this subject and further questions the route for the most ethical outcome, noting down statistics but also taking a critical perspective of how those are interpreted and collected. Finally, the academic journal acknowledges the possibility of the moderator interfering with the implemented framework referring to it as the possibility of 'private censoring' as it is according to the moderator's best interpretation of illegal hate speech. Due to the size of the platform Reddit and the level of dispersed interests that exist within the platform, moderators have a high-level of autonomy to moderate the content coming through on their subreddits, and therefore critically questioning the validity of their framework is necessary.

Criteria of Controversial Content - 2.3

In the absence of official law, scholars such as Meneses (2021) have attempted to concern themselves with controversial content. His work is important because he managed to set up a framework and process of controversial content identifiable by 4 criteria. Although not official or enforced, it sets up the precedent for solving the issue and becomes the groundwork for theory on the matter. The motivation of his work resides from the idea that there is a lack of critical thinking (CT) when it comes to consuming unverified content on the internet (Meneses, 2021, p.1). The explained crisis of this event was set in motion due to the COVID-19 pandemic, in which concern on this subject became one on the rise as a recent survey outlined that half the surveyed Canadians believed 1/4th COVID-19 myths (Romer & Jamison, 2020), more than 1/3rd of the US population believe in extreme explanations for the COVID-19 pandemic according to Pew Research (2021). Thus, despite the existing regulations on platforms as mentioned by EU and US law, harmful disinformation has proven to be effective.

The work by Meneses essentially builds on the theory of Mills (2011) extending the notion of the need of a scrutinized self-regulation "Digital media lends itself to this by allowing the dissemination of almost unlimited quantities of unverified content… …Learning how to evaluate online information is thus vital in such scenarios." (p.1). Meneses provides four criteria on evaluating online content concerned with controversial issues, Author position, Author Motivation, Systematicity and Scrutiny. Author position and Author motivation are deemed as 2 dimensions, the first made up of "aspects related to the author's knowledge quality, expertise, or reputation given by training/education, experience, specialization/relevance, publication…" (p.7) and the latter on aspects of transparency, which is weighted by honesty objectivity, bias and conflicts of interest (Meneses, 2021, p.8).

One of the more concerning characteristics of the process is that there is inference and judgement, and a tendency to generalize (Kahneman, 2011, p.85; Meneses, 2021, p.9) As it is the role of the moderator to encourage discussions and different perspectives, the judgement of the moderator is imperative to growing the discussion with then the expectation of understanding how to properly combat hate speech. Meneses introduces the "systematically or unsystematically" method, in which the latter serves to add more reliability to a claim (Meneses, 2021, p.9). Facione & Facione (2020) describe systematicity as a disposition of the mind, in which approaching problems is done in a systematic manner, therefore disciplined and orderly. Essentially referring to the research method "employed to collect, analyze, and synthesize" data to generate valid information or knowledge" (Facione & Facione, 2020, p.10; Facione & Gittens p.176). It assumes for example that the sample of the representatives by the

"collection, analysis and synthesis" of data is sufficient to confirm a generalization (Meneses, 2021, p.10).

The final criteria revolve around the concept of scrutiny and follows the logic of corroboration. Fischer (2011) implies it plays a fundamental role in the evaluation of information (p.98). Moreso, it has been deemed as "Civic Online Reasoning" (COR), 'skills' which focuses on lateral reading deemed as a process to check the trustworthiness of a source of information, through other independent reliable sources; essentially viewed as an imperative method of determining information accuracy (McGrew et al. 2018; Mason et al. 2014).

The Reddit Ecosystem – 2.4

Another macro perspective is on the platform itself and its eco system. This is where most motivators reside for the moderators, as they are directly linked to users, admins and other moderators on the platform. Thus any action or rule(s) which are implemented or changed are instant. There is no broker to consider and content must be directly removed or reviewed depending on subreddit functions. Within the ecosystem the admins have the highest motivators as no one can veto their decisions. What indirectly concerns the ecosystem are investors. Reddit as a private business can conduct it with any actor, in which they were held with wide critique to after accepting a 150million investment of media company Tencent. Despite assurances the investment would not affect their conduct, users doubted their honesty as Reddit already lacks transparency to their userbase and volunteer staff moderators (Potter, 2021).

If business is conducted this way, it undoubtedly brings up speculations. As discussed to how the profit structures are highly dependent on investors, the influence Tencent can exercise over Reddit raises concerns. Furthermore if Reddit is sincere in staying autonomous in their decisions, it means Tencent views other commodities to Reddit as highly valuable it could exploit, such as data. In their transparency report of 2021, Reddit declared themselves autonomous pointing to their mission statement. "Users can be themselves, learn about the world around them, and be entertained by the content created and shared by our global community" (Reddit, 2022). To enact such a mission statement, Reddit would need to actively place motivators to achieve that goal (Potter, 2021; Jhaver et al., 2019).

The problem Reddit faces is that it is perceived by the userbase differently and thus displays a fractured identity as a result from a 2012 interview by Forbes. Within that article, co-founder Alexis Ohanian stated that Reddit would always be "a bastion of free speech". This would indicate motivators for implicit moderation. This is also showcased on their mod support program, suggestions and website (Reddit, 2022). These tools however contradict with profit structure, as explicit moderation is easier to maintain for PR reasons. Any controversial content can be removed and any user objecting to it banned, removing the controversy and thus problem in a much faster process. This causes Reddit to fluctuate and fail to maintain a standard set of values, mission statement and overall goal due to their attempt at keeping up with the rapidly changing landscape of online media (Barr, 2019). Dubbed the digital media ecology, Barr (2019) explains how the digital media and information online rapidly changes without forewarning, leaving laws, corporations and platforms to catch up and regulate such matters.

The structure of Reddit has been described as a hive mentality by Mills (2011). This structure is called the positive feedback loop. This means that content is voted upon and those with a positive score are pushed to the front (Mills, 2011, p.4). The rise of such a process was necessary due to the quantity of information across the internet and the rise of the attention economy, it was essentially as a solution to information overload (Mills, 2011, p.4). This structure allows a people-powered approach to provide the most relevant information through a democratic method (Mills, 2011, p.1). However, the system is deemed vulnerable as the requirement to moderate content up to the discretion of the moderator allows them to choose what popular content is allowed and which is not. In the case a moderators fails to identify disinformation, such content is then likely to gain traction and be popularized and believed among the masses of the community (McGrew et al. 2018). Ideally, moderators are expected to hold a high level of media literacy, however due to the nature of the recruitment on Reddit as volunteers that frequent the community are prone to having their own biases contribute to their moderating decisions (Glenski & Weninger, 2017; Medvedev et al., 2019). Despite the suggestions by Reddit for implicit moderation on their moderation website, moderators are not required to follow those standards as they are not enforced (Mills, 2011, p.7; Achimescu & Chachev, 2021). The argument made for an increase in explicit moderation is the ratio of influx of content to moderators, in which the lack of a wide array of moderators causes for stricter moderation practices to keep up with the amount of content that is posted (Conner et al., 2007). The relevance of ethical decision-making begins to show through these scenarios, and although unpopular, strict moderation styles are possibly necessary.

Besides active moderation of content, according to Reddit, it is also the moderator's job to encourage discussion. They hold the power to ban a certain expression of posts for a given time, to either incentivize creativity and/or bring minority opinions to the forefront to be discussed, in order to prevent the creation of an echo chamber (Mills, 2018). However an overview of the website suggests, accusations consistently point out Reddit's inability to enforce these standards on many moderation teams of hugely influential subreddits which choose explicit moderation styles on opinions they disagree with rather than an equal playing field (Mills, 2018). This establishes a root that is deeper in Reddit rather than a moderation problem in itself.

In the most recent example, on the subreddit r/polls, the question of "Does Reddit have a political bias?" arose in which 613 users (82%) voted that Reddit have a 'politically left wing bias', with further comments indicating the obviousness of this theme, with the highest voted "Go on popular for a minute and you'll know", followed by "It's really strange, I think politically reddit has mostly leftist opinions but it's also really incelly" and "Overwhelmingly biased to the left. No argument.". The poll was eventually removed for violating a rule, in which users further speculated that moderators used it as an excuse to remove evidence of Reddit's political bias. This case does more than outline flaws of moderation by ethical standards, but instead also includes the idea that controversial content is evidently tied to the topics of politics. It must be noted that this case may not even be due to Reddit having a political bias, but comes as a result from the lack of transparency by the administrators on standards, that feed into conspiracies and further diluted the trust between moderators and users.

Rules – 2.4.1

Rules of a subreddit are formed by the creator of the subreddit and moderators hired for the subreddit are able to change them too if need be (Reddit, 2022). Besides the legal obligations of the website to remove illegal content, only the existing sitewide rules of the Reddit are further enforced. This is named the 'Reddit Content Policy', which stresses 3 points. The first, is post authentic content to subreddits the user is engaged with. The second, to not cheat or engage in content manipulation and third, to generally not interfere or disrupt Reddit communities (Reddit, 2022). Despite sitewide enforced rules, Reddit holds suggestions on how to create a proper online discussion forum and environment on their moderator help page.

The suggestions there stress a high level of transparency for a healthy community environment. The rules will set the expectations of the community, which means that the users visiting and those active in the subreddit will understand how they are enforced. This will also

give the opportunity for the moderator to point to something, when enforcing with the moderator tools (Reddit, 2022). Rules can be edited and added/removed at any time using the mod tool hub, under rules and regulations. The official rules outlined by the moderator will always be visible under the community rule's widget, which must include a detailed description of the what the rule entails.

Reddit emphasizes a section named 'Removal Reasons', in which it explains the importance of rapid communication to why a post was taken down. As a means to save time on the job of moderation, Removal Reasons serve as a function to send predefined reasons to users if they happen to break rule. These reasons often appear after a moderator has removed an item from the mod queue, spam queue and all remaining queues.

Content controls is another category emphasized by Reddit. Content controls set up official guidelines within the community, and can include the addition of title and post requirements. In this instance the user has to meet said requirements in order for their post to be approved and posted within the community. This tool aims to reduce post removals for well-intentioned users and lessen the content to moderate for moderators (Reddit, 2022). These tools a mod tool hub and can set up requirements such as required words, Ban words for title, post and body, require or ban links from a domain and restrict the number of times a link can be posted.

Subreddits – 2.5.2

Subreddits is what Reddit had mentioned as their focal point in their mission statement, and are the communities themselves. A single subreddit is a community on Reddit itself with a specific interest or topic in which people gather to discuss (Reddit, 2019). Users on the website are pushed to subscribe to topics of their interest, and essentially any one user can create a new subreddit, albeit under the rules of Reddit itself. Strikingly, there are concerns with the method Reddit employs to allow or ban subreddits, despite its already spotty reputation with moderating them (Mills, 2018). First, to be able to create a new subreddit as a user, you must register to the site. This is a low barrier of entry as the website does not even require for the user to register an email, therefore attracting many users due to the ease in which it is to sign up on Reddit (Ovadia, 2015; Mills, 2018).

Next, the user must name and describe the subreddit through the Reddit UI. Though among these standard processes the user must also acquire a certain level of 'positive karma' (Reddit, 2019; Ovadia 2015). Karma is the point system Reddit uses for users to either like or dislike a comment and/or post. If a post or comment is upvoted it gains one point. Thus the intention behind it is, to prevent spamming the creation of subreddits as well as having bots creating ones and in turn lessening the work to maintain subreddits (Reddit, 2019). However, in practice this comes off different to users as a flawed and exploitable system, in which users with very high levels of karma keep creating and maintaining subreddits of their interest and according to their rules.

Thus if a user with lower karma wants to create a subreddit with the same topic, but may have differing views or expectations, are unable to do so or gain traction enough to grow the community. This is due to high karma users having the ability to instantly out compete them by creating subreddits quicker, and by having posted more popular posts in other subreddits likely gain more followers, therefore when creating a subreddit notifying them and highly increasing the chances of creating a more popular subreddit overnight in contrast to a user who may be newer to Reddit (Ovadia, 2015).

Furthermore, many have concluded the bot argument invalid, as users simply use bots trained to post highly upvoted content and reposts on high volume and popular subreddits with a high voting rate to instantly gain a large number of karma and followers; therefore allowing them to create more subreddits according to their own rules and values as they desire (Ovadia, 2015). Finally, the concern is, is that mods with high levels of karma influence and uplift other users with their same values and ideas within their communities, essentially almost a close variation of nepotism, to a high karma user and erode those with different views and values. The problem herein lies with that it is done at the level of a moderator, the individual which can influence what information can be posted and commented within a community, potentially increasing the chances of it becoming a polarizing echo chamber (Geiß et al., 2021). It makes it all the more questionable for moderators to then engage within those subreddits with controversial content as they get to decide what is controversial based on their values, views and network rather than a universal, sitewide or democratic agreement. It is Reddit's unique focus and structure of subreddits that apply this specific moderator problem. This is not to say all subreddits ascribe to such a structure, but merely include such possibilities and exploits which have already occurred (Ovadia, 2015).

Moderators - 2.4.3

A moderator, officially defined by Oxford Languages (2022), is "a person who moderates an internet forum or online discussion. Wright (2009) describes the moderator as necessary, keeping the citizen engagement focused and ensures the contribution of value towards the online debate (Wright, 2009, p.2; Kearns et al., 2002). Yet this claim was written during the rise of the internet, before established services and structures were put in place and were formalized. Moderators may still uphold the mentioned values, goals and tasks, but may perform them differently as the environment and public sphere since that time has drastically been altered.

The role of the moderator has been investigated before. As users of SNS platforms aim build communities based on interests with each other to discuss them, the challenge within that setting is the moderation (Almerekhi et al., 2020, p.3). Moderation was a necessary role that evolved due to free-space communities exhibiting high amounts of 'toxicity' and 'negative emotion language'. Within these spaces, comments of users continuously endangered the discussions. E.g. "b!tch were not not talking about that so I hope you get f!cking r!ped u f!cking whore", (Almerekhi et al., 2020, p.1).Pew Research has stated that in 2017, 41% of Americans were victims of online harassment and 66% witnessed harassment behavior at others (Almerekhi et al., 2020, p.1).In response to these numbers, online services have understood the need to fight harassment and cyberbullying and have done this through setting up clearer guidelines and even legally enforced sitewide rules (Almerekhi et al., 2020, p.3; De Streel et al., 2020, p.30). To enforce these set of rules, a system has been set in place including human and ai moderators to further fight ongoing toxicity within these spaces (Almerekhi et al., 2020, p.1).

A short overview by Jasser et al. (2021) and Almerekhi et al. (2020) reveals that moderators on Reddit are given the tools to censor, ban, shadow ban, remove comments, remove discussion threads, pin posts and more, to influence the subreddit. It furthermore provides them with the opportunity to influence the botted AI as it has to adhere to subreddit rules and nuances which puts into question how moderators go about practicing their ethical framework, or justifying it. The goal of the system is to ultimately remove toxic posts and penalize users who have engaged in such practices (Almerekhi et al., 2020, p.1). Moderation includes the removal of or keeping content users may personally in the community disagree with. This disagreement can range to the extent of advocating for the removal of said offensive content, or keeping it as a means for critical discussion and education. The ultimate decision is of course conducted by the moderator who dictates the rules, and exceptions to those rules

depending on circumstances, hence the importance to their ethical-decision framework. To discover this, more in-depth concepts have to be overviewed.

Moderation processes - 2.4.4

When decisions about a user's content occurs, it often appears through a lens of a 'Moderation Process'. This concept refers to the process moderators take, either explicit or implicit to conduct their decisions on content posted by users. Seering et al., (2019) proposed 3 processes in which moderators evolve their engagement with the community and fulfill their role. These include, actions, responses and tasks assigned to them alongside the rules that come with the growth of the community and in this instance, subreddit. The study observed these processes through algorithmic and non-algorithmic tools. Scholars had largely looked at "the effect of displaying moderation rules on the perceived bias in news… ...and the spread of harassment in online discussions" (Almerekhi et al., 2020, p.3). These studies had already shown that moderation intervention increased positive participation and lowered bias in the perception of science-related news (Seering et al., 2019). Another study done on moderation processes by Gibson (2019) focused on subreddits with the 'safe-space' label in which moderators of said communities had a higher rate of removing content and incited a higher rate of self-censorship. These types of communities tended to have a higher focus on positive emotion language (Gibson, 2019, p.1). Free space communities, with less moderation intervention, had higher rates of negative emotion language (Gibson, 2019, p.1). Yet the findings suggested it is integral to remain flexible and nuanced as the moderator yet it was also stated that different governance rules across subreddits are essential to maintaining an online democracy.

Next to moderation processes is moderation style. In a study done by Matzat & Rooks (2014), moderation processes led to 2 mutually exclusive direct and indirect moderation styles. This study was conducted on a Yahoo! Health forum, although not Reddit, is similar in structure and culture. According to the results, users preferred an indirect, incentive-based moderation style that leads towards positive interactions. Users tended to have a very negative view on direct moderation, with the highest aversion to penalization of contribution of content. Within a similar notion, Lander (2015) found that implicit moderation strategies were welcomed as opposed to explicit exercises of moderation and practices. Part of the moderation that occurs on Reddit is most often done by auto-moderators. It is important to understand that these auto-moderation, also named filter bots are implemented and created by moderators (Reddit, 2022). Thus their nuance and decision making is extended by these devices that dictate

what applies to the rules of the community and which content does not. These devices are used due to the the scale of some communities exceeding a million users on Reddit (Reddit, 2022). These filter bots are entities designed as filters for bad words and expressions, which are often aimed at comments. This necessity becomes apparent as when communities grow, so do the *toxic comments* (Almerekhi et al., 2020, p.5). Toxic comments include elements of profanity, harassment and hate speech. Following moderation processes pushes upon the notion of how much they are able to achieve with the tools at their disposal. This prompts a look into moderator tools and the possible scope of their power.

Moderator Tools - 2.4.5

To fully understand the perspective of the moderator an in-depth exploration of their functions has to be conducted. Understanding the hands-on functions of moderators can indicate their ability to filter information on the website. These functions are all displayed on the official Reddit sister website, mod.reddithelp.com. The website depicts a total of 10 categories for moderation tools. These include, Moderation queues, User management, Flair & Emojis, Rules and Regulations, Content, Other, Modmail, Chat Settings, Community Activity and Mod Log.

The relevant categories regarding the decision making on content are separated into 3. The first, Moderation Queues as it refers to managing user reports, spam, edited content and unmoderated content. Second, User management, this refers to the banning and muting of users, adding approved users and managing moderators and mod permissions. Third are rules and regulations, which refers to how rules are defined within the community, the reasons for post removals, implementing guidelines to aid Redditors in following community rules and how to use the Automod.

Moderation Queues - 2.4.6

Moderation queues primarily function as a central listing of all the content in the community a moderator has to review. These includes, user reports, filtered posts and comments. Notably this is separate from the spam queue, in which filtered content is directly sent to separate inbox. The modqueue provides 4 actions for the moderator to engage in. Approve, which approves the selected content and makes it visible if it was previously removed as spam or filtered. Remove, which removes the content from the community, although it is still visible within the spam folder thereafter. Spam, will list the content as a spam item, which as a decision is fed into the algorithm to learn what items tend to be spam and which are not.

Ignore reports, in which a piece of content may be reported several times, but is not breaking community or sitewide guidelines, in which it ignores future reports on the content. Among the decision to remove a piece of content, there is the added option to add a removal reason, in which it brings attention of the community to why a piece of content was removed, as to further reinforce and educate the users about community guidelines.

Reports - 2.4.7

Although similar to modqueues, another category for moderating content derives from Reports. Reports are a result from posts and comments flagged through user-reporting. Moderators can also choose to snooze users, in which within a given timeframe the reports from a user will be ignored by the system and turned off for that moderator. After the given time is up, reports will be noticeable and available to the moderator again. There is also an option to unsnoozed a user if the moderator desires to engage with the reports earlier. The spam folder consists of removed content only. The moderator can enter the folder to reinstate content that may have accidentally been removed. Edited queue only contains content which has been edited. Finally, unmoderated queue lists all submitted posts by users that have yet to be moderated or revied. Anything posted within the community will appear in this queue, in which it still needs to be approved, removed or marked as spam.

The User Management category concerns itself with moderating users within the subreddit. According to the Reddit guidelines the intention of these tools is only to be used if educating a member of the community on the guidelines first and foremost has failed. Reddit issues this tool to moderators in order to keep the peace. The moderator has thus the option to ban or mute a user. The option of banning a user prevents them from posting or commenting within the community, either indefinitely or for a limited period of time. When banning a user the option to give a notification with a reason is included, in which the moderator must outline what rules have been violated and to limit confusion, follow-up messages or repeated offenses. Muting is primarily used as an add-on after the ban if the user continues to spam mail and messages to the moderator. Reddit itself advises banning and muting to be used sparingly and if stated reasons are clear and transparent in line with expectations of the community, as it is the healthiest approach to the Reddit public sphere (Reddit, 2022).

User perception on Moderation - 2.5

Observing the relationship between the user and the moderator is an important aspect to take note of in Reddit moderation. Unlike moderators of other SNS platforms, there is a

more direct relationship between the user and the moderator. This is because the recruiting pool mainly consists of existing active users of the subreddit. The main reasons include their understanding of the culture within the subreddit and most users associating that individual with an already positive reputation, making moderation efforts smoother with less disputes (Almerekhi et al., 2020). Almerkhi et al., (2020), however also expands upon how it may increase nepotism and rule bending within the community, in which users are treated unequally due to some having personal relations with the appointed moderator. These flaws become highly accessible as Reddit does not have a strict policy to attend to regarding moderator exercise of power (Almerekhi et al., 2020, p.6).

The paper by Almerekhi et al. (2020) performs a quantitative method through statistical modelling, investigating toxicity triggers within Reddit discussion threads and heavily refers to the element of moderation within the paper to help detect and remove such threads. Within the study itself, it revealed through an online survey, in which harassment against moderators included six independent variables. These six variables were linked to an increase in harassment activity across reddit subreddits. The study further states how 2 of these variables exist as a result from moderators struggling to manage the safety of online spaces for discussions effectively, often due as a result from the size of the subreddit, influx of comments on a thread and the overall complexity with dealing with toxicity. Almerekhi (2020) proposes a solution to ease the task of moderation through automated detection of toxicity by "finding the root" within the discussion thread by observing comments (p.2). This concept is coined as toxicity triggers. Additionally, it has to be recognized that toxicity triggers need to applied differently according to community culture as nuance is integral to judgement of the toxicity.

Two important variables are the responses of the users towards moderation practices of their content and moderation style (Almerkhi et al, 2020, p.2). There exist users who act aggressively towards moderators in these scenarios, and can even lead to targeted harassment (Almerkhi et al, 2020, p.2). Furthermore, as mentioned above, this paper took an in-depth approach at furthering the study of how strict moderation has a negative relationship with user experience and their ability to enjoy engaging discussions (Almerkhi et al, 2020, p.3). The notion that was supported that came from the study was that the ultimate goal of the moderator, as seen by the user, is to support the user and their freedom of speech without limiting their expressiveness. Hence the gatekeeper to a successful online democracy, as the space of Reddit is perceived. With the variety of variables in play of how moderation is perceived by the user, and the user experience correlated to moderation style, the goal of balancing discussion enters a grey area. To further establish the perspective these variables, arise from, the overview of

how controversial content is defined is the next step to properly understanding the full framework.

Chapter 3

Research Methodology – 3.0

The research method is comprised of the methods used to answer the research question. For this research question, the method qualitative in-depth interviews was most appropriate. This method of interviewing will also be semi-structured as to maximize meaning-making during the process. This requires the researcher to ask open-ended questions, as well as follow-up questions during the interview as to incentivize a combined effort to gain the maximum meaning in the given time frame (45-60 minutes). Then the, results will be conducted through thematic analysis.

Interview Method – 3.1

The interview method that will be followed is that of Johnson (2011), that specifies on semi-structured interviewing and how to maximize meaning-making. The below explained process will be followed as closely as possible during the interview procedure. First, Johnson (2011) explains how the preferred methods of in-depth interviewing is to conduct it in-person, face to face. This increases the level of intimacy and causes for the effect of acquiring a deeper understanding of the information and knowledge on their perception. If trying to follow the theory explained by Kendall (2008) and understanding nuance and ideology is quintessential to achieve the deepest understanding, in person interviews are the preferred option with Moderators. Furthermore, an in person real life setting is supposedly a more comfortable space for the participant to share information about him or herself according Johnson, 2011. However it is vital to note, this paper represented the method in 2011 in which the internet was still in its developing era. As Moderators spend several hours moderating content on the online, they naturally spend much more time online and thus this notion may be the opposite. Thus in the case of in-person interviews not being possible or the alternative is preferred, an online face-to-face, or a simple call through social platforms they moderate on such as discord or Reddit can be arranged. As moderators are hard to reach as a sample group, due to the mention of online targeted harassment, (Jhaver et al., 2018) the researcher must remain flexible with the given requests of the moderator once contact is established.

Sampling Strategy - 3.2

To gain a proper understanding of the space, topic and current affairs, and in the interest of time within research, 10 moderators will be sampled. Contact will be established by visiting

the platform of Reddit itself and using its functions to contact moderators. Not only does the platform display the names of the moderators you can choose to message, but it also allows for various contact methods to reach out to them, increasing the odds of responses. This includes, Message the mods button, Reddit chat and Direct Messaging.

The selection of the subreddits will include those of gaming subreddits, watch dog subreddits and internet phenomena subreddits. This is because on average those subreddits typically have the highest response rate whilst maintaining a large subscriber count (over ninety-thousand) (Reddit, 2022).

Messaging the moderators is also a requirement and a formal message is preferred to be constructed as to keep in mind the weariness moderators may concern themselves with regarding abuse they may often face (Almerkhi et al., 2020, p.9). The goal is to at least acquire 15 responses, and if it is not possible snowball sampling will be applied to increase response rate. This means that the first participants will be asked to refer to other potential participants in their network. This is especially effective for this space because acquiring a moderator's trust in the researcher will likely not only increase response rate but aid in the meaning making process of the interview itself. At last before the interview is conducted, the moderators must fit the sampling criteria which means they are and have actively moderated on Reddit for, at least, 4 months on subreddits with over ten-thousand subscribers with topics related to gaming, watch dogs and internet phenomenon's. Again, criteria remains flexible to maximize response rate, hindered by the previously mentioned obstacles.

Interview Guide – 3.3

In order to gain more out of the data that is being gathered, the researcher is advised to create an interview guide (Johnson, 2011; Showkat & Parveen, 2017). An interview guide is neither a time-consuming task when research such as in the literature review has already been conducted. Essentially, the researcher has to apply the knowledge of the theories and transform them into potential questions for the participant to answer. For semi-structured interviews, open ended questions provides the most opportunity for the participant to apply their own values, beliefs and perspective into the answer (Johnson, 2011).

Finally the interview guide also helps to correctly conduct the method of analysis. As the paper by Braun & Clarke (2012) suggests that because the thematic analysis method is used for "Systematically identifying, organizing and offering insight into patterns of meaning" the guide gives a direct focus and direction to themes, which occur within the given dataset (p.55-57). This method is a flexible approach which gives the opportunity for the research to conduct

both a deductive and inductive style which is integral to understanding perceptions and feelings that occur for moderators when dealing with content to the related topics of internet culture, controversial content and insincere moderation. Maxwell (2018) states, "concepts, assumptions, expectations, beliefs, and theories that supports the research" (p.2), indicating how it plays a major role in assessing and discussing what themes come to play during the interview.

Furthermore, it ensures the research author is knowledgeable on the latest themes and thus when replies are given by the participants, it would deliver a deeper insight on the subject. According to Barriball, (1994), the quality of the data can largely be affected by the type of interview guide and method the researcher decides to employ. In this instance the researcher will follow a semi-structured interview, meaning the interview guide is flexible and malleable depending on how the questions are answered. Thus, open-end questions will be implemented as they allow the participant to lead the discussion into what they believe is important, by which the researcher can follow and ask contextual meaningful follow-up questions (Galletta, 2013). Next, the semi-structured interview provides incentive for the participant to more freely express their opinion, therefore referring to the inductive approach (Galletta, 2013). As a result, both the participant and the research author engage in the meaning-making process, as a means to explore the existing or newly discovered themes (Barriball, 1994).

Tools – 3.4

One of the hands-on processes for the researcher are the tools. Tools help immensely in the structure of the research method as they are the core physical tenants to storing and processing the data. Any use of the tools related to conducting the interview with the moderator are mentioned before, or at the start of the interview to maximize transparency as for ethical responsibilities as a well as a show of good faith to maintain trust. The programs included during this research method include Voice Recorder, Descript and Microsoft Excel. The processes of using these apps are also described as to ease the replication of this method. Though variations of the apps are permitted if research is conducted this way as long as the quality remains the same and they achieve the same ends.

For during the interview, a phone recorder device is used. Within this research it was an iPhone XR, installed with the app 'Voice Recorder'. The app was downloaded via the Apple app store. Each interview was conducted through its 'record' feature, which can be found by tapping the main display which is a microphone symbol at the centre bottom of the app. During the interview it is possible to manipulate the audio in case unexpected events or interferences

occur to disrupt the interview. This includes a pause button, and a resume record button. As the app can record and store audio, it also has several routes to upload and retrieve the recorded data to a cloud or a laptop device. The function that will mainly be used is that of the WiFi sync, which gives a passcode from the app to an internet browser window, allowing only the researcher to access and download them for the time being and as long as the app is open, ensuring the protection and privacy of the data and the participants.

Next the program Descript is used for transcription. First, the downloaded interview audio file will be imported into the program. Next the program gives the option to convert the audio file into a written transcript which will be selected. After automatic transcription the program also has the option to automatically detect speakers, which assigns the aligned texts to the two lables, the interviewer and interviewee.

The researcher then has to manually check and correct the mistakes. The program includes several features, providing features for the most accurate transcription. This includes an audio log to listen along the highlighted words being said when checking the transcription, the ability to edit the transcript and an audio speed button to slow down the audio in the case a sentence, phrase or word is too fast to comprehend. The effect works vice versa and the speed can be upped in order to move through the transcription faster, to finish the process of transcribing faster. During this period, the researcher can also start part of the operationalization, of finding open codes which will be explained in the following section. Finally, the transcription can be exported through the program function of publishing, in which it will be converted to a pdf or word file. During this process the feature ‘thirty-second interval’ is selected to provide additional time stamps within the transcripts as a tool to organize the written transcript by timing.

After transcription the data will be coded through highlighting the transcripts in different colours, essentially colour coding the process. A table will be set up separately to define which theme is related to which colour.

Analysis Method – 3.5

One of the most efficient methods to analyse meaning making results is through Thematic analysis. Thematic analysis consists of observing patterns in the data through codes the researcher creates along the way of analysing the outcome of qualitative interviews (Braun & Clarke, 2008, p.6). This process of coding is both efficient and effective to analyse full interview transcripts in a restricted time amount, allowing the researcher to put their focus on interpreting the results and thus synergizing the codes with meaningful discussion that occurs after the results have been finalized.

In order to execute the process successfully 6 steps have to be taken. The first requires the researcher to familiarize with the data. This occurs during the interview. After, the data collected is transcribed to written form. Despite this method being favourable to time management, this step remains the most time-consuming and tedious. Yet, it is the best method to acquire the most detailed form, including body language, facial reactions, tone and punctuations.

The next step consists of building 'initial codes'. This happens in tandem with familiarization in which the researcher composes a list of ideas on why the data is relevant and/or what makes it interesting. This can be achieved by noting down sentences, phrases, quotes or summaries of the analysed text. The researcher thus thoroughly analyses the entirety of the transcripts, and finds parts to systematically code in reference to a feature, segment or element that can be assessed and generated into a meaningful phenomenon (Boyatzis, 1998, p.65; Braun & Clarke, 2008, p.16).

Next the third step is related to finding themes, which can be made up of several codes. These codes may form a pattern and those can then be organized into categories of 'overarching themes' (Braun & Clarke, 2008, p.92). Thus, the researcher is required to study the themes in detail in order to create, three forms of themes, including the one mentioned above, and both 'sub-themes' and 'miscellaneous themes'. Miscellaneous themes refer to codes which are not recurring but may be notable information that is important to meaning making. Finally, the creation of a visual thematic map is a suggested method as it organizes the data and makes it easier for the researcher to analyze and connect all overarching themes and sub-themes (Braun & Clarke, 2008, p.99).

Following, is the fourth step, which resides in reviewing the chosen themes and sub-themes. The researcher has to carefully look over and determine if the patterns of the codes are comprehensible and logical (Braun & Clarke, 2008, p.101). During this process the validity of

the themes are compared alongside the theories discussed within the theoretical framework in order to build a reputable and sound outcome.

The fifth step is to label and defining the themes, by which the definition of the theme should reflect the features, patterns and aspects that was processed within the previous phase (Braun & Clarke, 2008, p.101; Boyatzis, 1998). Finally, the sixth step consists of producing the report in order to tell the narrative that was formed from the patterns in the data (Braun & Clarke, 2008). The data must be portrayed "in a way which convinces the reader of the merit and validity of your analysis" (Braun & Clarke, 2008; p. 93). Finally, a 15-point checklist written by Braun & Clarke (2008) will be utilized in order to check if the thematic analysis was appropriately applied.

Chapter 4

Results and discussion – 4.0

After the processing of the interviews of the 10 moderators, and coding has been conducted the results are posted on the appendices of the thesis. The notable outcomes of the results are the 10 selective themes, also displayed within the appendix (figure ?) as Moderation Style, Nuance, Interpersonal Relationships, Controversial Content, Human-Bot Collaboration Moderation, Ethical and Philosophical framework, Reddit ideology, Reddit Moderator Culture, Insincerity and Social Commentary. Finally the sub-themes function to complement the existing themes, and although not the main focus, still serve a purpose of displaying important features that come with online moderation of content. All the results have been displayed in tables to provide clarity of the coding process that occurred in the previous step ranging from initial/open, axial and selective codes to the colours used per code.

Before discussing the themes, the first result to take notice of is how 8 out of the 10 interviewees applied voice changers/scramblers during their interviews, adamant to keep themselves as anonymous as possible. When asked about the use of voice changers, the main response was how they feared possible repercussions of Reddit staff/admins they were already at odds with. Despite repeated assurances of their privacy and ethical rights of the research, voice changers was a necessity to them. Another notable conduct of attaining interviewees during the process was how snowball sampling became a requirement due to the mentioned weariness Reddit moderators may display due to harassment. The surprising output of this was that the theory stated how weariness would result from user directed harassment, but instead it was clear this was not the only group to worry about as again, it was the weariness of repercussions from Reddit admins that became the focal point of this worry.

Finally, the 10 respondents which participated derived from subreddits of different genres, which included Gaming, (r/Mordekaisermains, r/Stellaris and r/GhostofTsushima), Entertainment related to movies and shows (r/AoT), Watchdogs (r/WRD), Comedy (r/Greentext), Sci-fi entertainment (r/Warhammer40k) and Conteroversial (r/ChrisChanSonichu).

Moderation Style – 4.1

The theme of Moderation style features the data which explains how content is moderated on the platform. This includes the styles mentioned by Matias (2019) figuring whether the moderator is implicit or explicit, or a mix of both in their moderation. This theme is distinguished however between the actual practice of moderation and the preferred style of

moderation. This theme highlights the hands-on real time practices the moderators implement regardless of the their practice preference. As moderation style is not only up to the discretion of the moderator and is also enforced by admins, laws and regulations. The goal of this patterned theme is to understand and distinguish the real-life implemented practices from the ideological preferences of the moderator themselves, or how they choose to apply their ideology within the moderating constraints, if at all.

Moderation Style is the first and one of the most prominent themes featured that remained strongly linked to the theory. This is because every moderator had detailed input on their moderation practices which was directly related to implicit and explicit practices. The theme of moderation style became a subject of debate of preference versus necessity. Every single moderator described their moderation style, but also expressed how it was not how they preferred method to go about the practice. First, All 10 participants expressed how their moderation style was largely situational and mixed between implicit and explicit, after the question was asked which style was more akin to them. However without further provocation on the question, 8 out of the 10 respondents immediately implied that this was not entirely their choice. This is because Reddit admins hold a large input on the moderation style of a subreddit if they please. Some subreddits also have third-parties they are affiliated in which their topic is linked to other official institutions that put restrictions in place, e.g. r/Stellaris moderator is linked to r/ParadoxPlaza, that hold real staff from the studio game developer enforced restrictions leaks and unverified patches of the game (interviewee 6, moderator of r/Stellaris). Surprisingly, legal aspects were mentioned very rarely, with the theme appearing only 6 times in all 10 interviews, as the sub-theme suggests.

When further discussing this the theme of moderation style, for preference, it was clear that implicit moderation was favoured amongst 9 of the 10, respondents. However the 1 respondent that favoured an explicit approach is moderator of subreddit that revolves around a very controversial individual that requires more reviews of content before posting is permitted. Essentially only a explicit approach can be applied to his case because the topic requires it. The implicit approach is mainly seen as a method to keep the flow of discussion, with one respondent (interviewee 1) voicing its use to, “obviously, uh, prevent disruption to the ordinary flow of the subreddit”. This favour goes as far as giving users the benefit of the doubt and providing several warnings instead of censoring or removing their content, as mentioned by interviewees, 4,5,6,7 and 8. According to the respondents, as the theory suggested, explicit moderation is frowned upon by users, and even most moderators. The exceptions that were explained to this however, were of larger subreddit sizes in which decisions about content had

to be made much quicker to be able to keep up with content volume. The majority, 8 moderators admitted they were likely to switch to a more explicit style of moderation as well as add more automated systems with explicit functions if their subreddit were to grow exponentially. “Yes. It gets the point where there's too much for ordinary humans to handle. You can't go through every thread, every single conversation chain and post reminders or auto, like correct.” (Interviewee 1), “I think that it really depends on a handful of things. The first issue has to do with the size of the forum…” (Interviewee 7).

Finally, another aspect that highly impacted moderation style to be more explicit and also defined as a sub-theme is rules. Rules of the subreddit are the clearest example and reasoning to how a moderator approaches their job. The first focus is checking if the involved user has broken any rule organized by the subreddit, and if so, if under no exceptional circumstances would be explicitly moderated and removed. As Interviewee 5 explained, explicit moderation is justified when the rules are transparent. Rules exist mainly to keep the subreddit on ‘task’, and an incentive rather than a reason to actively exercise explicit moderation (interviewee 2). Thus the interviewees explain, how it becomes concerning when explicit moderation is actively used, even when unnecessary and attribute it more to disingenuous moderators with ulterior motives other than to moderate for the topic of the subreddit. These concerns of will be discussed in a later theme, referring to Incinserity.

Nuance – 4.2

The theme of ‘Nuance’ refers to small and subtle differences in the manner in which content is understood, this also applies to the online space. Culture within a subreddit is highly valued, but is just one of the several expressions nuance presents itself in regarding that matter. As the topic of a subreddit is the main reason for its existence, nuance is what is understood as the subtle understanding of the topic. This means the individual critically understands the content they’re engaging with and can therefore better understand how approach the topic and have a more educated and valued opinion to contribute to dealing with online content. This includes and lived experiences, interpretations and perception of online content and internet culture.

The results of the theme nuance revealed how moderators observed the shift into a culture of intolerance towards the diversity of thought in mainstream spaces of Reddit. The main concern for the participants further revealed the fear culture that arose from this placing worries that those which do not understand their subreddit culture, and therefore nuance would be offended and advocate for the shutdown of their space. Most subreddits are spaces of topics for

passions and hobbies, which for good reasons highly likely only attracts certain users which are invested in the topic and show the same passion moderators do.

In all 10 interviews, the moderators advocated for a structural change to require moderators to have to be an active member of the subreddit who frequents the space or otherwise the lack of understanding would lead to unwanted and inappropriate rules that fracture the flow of content or even lead to an eventual ban if moderators and users of the space don't comply. This occurred with a subreddit named r/2balkan4you, as described by Interviewee 2. This case refers to a moderator admin of Reddit from the mainstream that found the content of that subreddit hateful and forced themselves as a moderator on the subreddit to manipulate the rules, as posts were mainly insulting, stereotyping and humiliating Balkan countries and their culture. However the subreddit turned out to be a space in which users from the Balkans posted this content as satire, due to the cultural implication of the Balkans historically always conflicting and enduring wars between each other. This critical misinterpretation of the Reddit moderator admin of the content for forcing changes was met with criticisms from users and moderators of the subreddit alike, which the admin then deemed as targeted harassment believing this was a result from the userbase being toxic, rather than upset for not understanding the deeper nuance that resided in the subreddit's culture. As a result the admin banned the subreddit, removing its existence and history off the website. It was later discovered that the admin was the only user on subreddit that was not Balkan themselves, which likely was one of the main causes to the misinterpretation of the content.

Interviewee 2, who mentioned the case within the interview, along several other controversial events that occurred on Reddit. This is because the user is part of a moderating group that archives scandals, double standards and censorship r/WatchRedditdie (or WRD), a position that puts the individual directly at odds with the admins of the website. Thus the answer must be observed with an understanding that biases to reporting of that case may be involved. It is notable to mention the subreddit as of writing this paper, has been unwillingly closed on 15/06/2022. Ultimately, all participants have expressed fear that such an event would occur to them too as a result from the lack of understanding meme culture, internet culture and hobby niches. Interviewee 6, moderator of r/attackontitan, a Japanese anime series and manga, expressed how some discussions may involve the topic genocide and several users supporting it within the series. Yet clearly provided context and culture on how such a position came about, citing that its users justifying the feelings and actions of a villain character, rather than the actual task. This is depicted in the quote "but because of the. way the story is this like egh... you kinda the have to say, yeah, that's fine. It's like people rooting for a side character, like

Darth Vader you can't stop someone from rooting for a villain even if they did something really bad,". All 10 interviews reveal that context is widely necessary and the only manner to achieve it is to personally already be part of the community. The blame is put to the structure of Reddit as interviewee, 3, 5, 7 and 8, all express the hyperfocus on Reddit for profit. This focus results in the lack of care for nuance, as long as it falls in line to make image of Reddit appealing to shareholders, stockholders and the mainstream.

(Interviewee 3) "Yeah. Yeah. That biases is like and it can be either way or it could be just related to a certain agenda or a way of seeing things like a bias towards, advertising, prompting, primarily advertising for profit.". (Interviewee 5) "There's always people that want achieve a certain end, I think and unfortunately, it's those people that ultimately, I think make more of a difference, uh, that make more of like a change in like the direction of the platform. It's nothing I can do about that. Yeah. Um, but I think. That the people. On top of Reddit, the actual management, um, I doubt that they really cared much, um, because Reddit already has like stockholders on the board and they don't really, if they're making a profit." (Interviewee 8) "You know, they want, it's basically like, you know, these things, they are businesses. They are trying to make money. And while every decision they make is not going to be the right decision or even the most profitable one, um, their main goal is to appease their stockholders or the shareholders or whatever the official term is. Um, or w or even just the owners, like it's a privately traded company. And if a certain, if they deem a certain, a person or a certain group of people to. Uh, yeah, too expensive relative to the amount of money they directly or indirectly bring in. Then they're probably going to be on the ground, penalize them in some way, shape or form like it's this is how it is, uh,"

Finally, interviewee 5 takes a wider approach and supports the theory of the information age and attention economy and results in nuance in being overlooked. Clickbait culture takes precedent, and if the subreddit becomes popular enough to come to mainstream page of Reddit, r/Popular, inevitably attracts audiences that will find something controversial or offensive and pressure the subreddit into changing rather than taking the time to understand the content more critically. This lack of nuance eventually puts the position of the subreddit in danger to be removed or replaced all together for a more generally acceptable and politically correct version. Lacking context and thus nuance is a widely debated topic outside Reddit itself and is a problem presented in various forms, the equivalent of which can be the attributed to the current debate around offensive jokes in comedy that have erected their own controversies such as Dave Chappelle and Ricky Gervais on transgender people jokes, and comedic actor Rowan Atkinson

most recently commenting on the lack of tolerance, understanding and context to humour (Pierce, 2022; Stolworthy, 2022).

However this is portrayed on several different levels, by different interviewees. Notably, interviewee 2, which described themselves as a "Free speech absolutist", described how certain derogatory sounding statements may have been from a position of social commentary whereby they state "So the US and Europe flagged this as hate speech and individual might express something like, oh, immigrants are ruining a country, or we can't have all these foreign peoples coming in. Um, to me, that's kind of brave. I, I think that could be interpreted as a statement about a country's political policies and about. Sociological problems that come up with an influx of human beings, not necessarily a dehumanizing criticism of those migrants themselves.". On the other hand, 7 other respondents out of the 10, clearly expressed that they drew the line of such nuance much earlier than at that statement, showing varying levels of tolerance to similar speech that is stated above by interviewee 2. This shows despite the unified agreement on the lack of nuance plaguing Reddit, moderators are still divided in opinions on where to extend the benefit of doubt, even with an implicit preference approach to moderation; hence causing the legal question of where to draw the line on hate speech to remain unresolved.

Interpersonal Relationships – 4.3

The theme of Interpersonal relationships in the Reddit ecosystem refers to the different links between actors and the meaning behind these relations. Meaning is constructed out of how the moderators view the different actors they interact with. Every stakeholder plays a part to defining the topic and experience of a subreddit and impact the modus operandi of the entire system as a whole. The sub-theme for the matter, third-party relationships typically have the same effect, but aren't applicable to all scenarios, and only impact isolated spaces.

This theme also reveals who holds the favour in the balance of power, and can typically reveal the attitude and motivations of individuals towards each other. One of the first implications of this is how moderators can easily hand over moderator privileges to any user. This is evident as all 10 moderators received their position as a moderator from another moderator. It was also a voluntary process, making all the participants volunteers. Furthermore, the majority of participants implied through their language use, that they were a previous regular visitor of the subreddit. This is evident from the sub-theme 'recruitment'. Interviewee 10 "It was open so I thought to myself, like I already frequently interact with the community. Like more than probably the average user. So I decided why not take care of it step further and I've begun moderating.", or interviewee 3 "Uh, I just found A subreddit I'm active, active on

and decided that, you know. Uh, if I'm going to be this active, I might as well apply for the position of a moderator"

Despite, this structure in place, several participants, 3, 6, 7,8 and 10, believed this notion should neither be extended to mainstream subreddits and nor should its process endure, official vetting processes have to be added. This referred to the sub-theme of structural change in which 2 participants talked about the structural changes they would like to see to reduce nepotism and increase integrity. Yet, both participants also had a pessimistic view on this as after their suggestions they knew the rigid system of Reddit was unlikely to change because their model seems so profitable. Due to moderators volunteering, no costs are spent on upkeeping them. Having spaces in place to motivate people for the position and keeping them unpaid, means compromising and allowing moderators more freedom with the tools and decision-making processes in order to keep them content. Yet, it also increases the chances of bad actors in those positions. For example activists or disingenuous types.

All 10 moderators also described their subreddit as outside the mainstream, hence their belief of informal hiring process applies differently to them due to trust relations and the intensity of the topic. Gaming and entertainment for example, described through various sentences that they were low intensity, and were unlikely to face content that would increase moderation tasks. Also, 6 of the interviewees believed, due to their niche subreddit topics, only the correct passionate people with the same values that understand the culture were to be selected for the position. 2 experiences include that over interviewee the only information available on those mainstream large subreddits and their interpersonal relationships on the site are the lived experiences and anecdotes these moderators voice through the distinguishing features of the different subreddits.

The first notable result is how the subreddit topic highly impacts the relationship the moderator has with other actors. For all 10 participants, whenever expressing details about their contact with other actors, would also follow-up indirectly relating to emphasis on subreddit culture. For example, those which participate regularly in the subreddit naturally would have a deeper understanding of the subreddit culture which was displayed as "Yes, there were always something there's never been a time where. Like, where you know, I go through a period where there's absolutely nothing. There's always, normally at least one bad actor. No means like the regulars. They just have a kind of a flawless track record because obviously the people that interact regularly are not the ones that are going to be causing problems." (Interviewee 1, r/Mordekaisermains, in reference to if controversial content is posted on the subreddit by users).

In other instances, there is confidence in the userbase between the moderator and user, with expectations the user understands the culture, which is available as internet culture comedy is a less niche topic. Due to the moderators wanting to keep it a expressive space, Interviewee 3 implies there is a level of blind trust for users, giving them access to more Reddit tools to manipulate posts, “Um, obviously it's kind of hard because everything is, you know, not safe for work. I mean, we go through a process of letting the users obviously mark themselves.” (Interviewee 3, r/Greentext). Thus to maintain the space in the manner they desire, with comedy available in any form and to efficiently be able to moderate the influx of all posts, there must be a unspoken understanding and etiquette users must follow.

Interviewee 3, depicted a stance in which the user held the power of a subreddit and they take a more democratic process. Moderators are the ones to usually finalize a decision, but it is also often backed by user reasoning and support. If a user is considered a bad actor, and acts unethically, there is a form of ostracization that occurs where users downvote and reply en masse to criticize their behaviour that can result in said user deleting their comment or account before the moderators have to make any decisions. As Interviewee 3 explains, “Um, I think of all of the times we're in the background, we're not noticed that much compared to other subreddits.”

Other results revealed that Reddit admins had an overwhelmingly negative presence on the platform, and that these were often users that are wildly unequipped to understand and handle standards, rules and expectations for subreddits, as the previous chapter had shown and as ties between moderators and admins reveals in this theme. This is because admins take a more top-down authoritative approach that protects the interests of shareholders and thus profit the structure. As long as profits are unaffected, the admins can do as they please without repercussion, whether it is enforce a political agenda, implement their own means in the website or turn a profit off content advertising.

Interviewee 6 voiced “but in terms of the management, they don't care. And I think terms are on the subreddit. It's too structurally broken for it to change it to any sort of productive direction.”. Interviewee 5, goes further into distrusting their fellow moderators of mainstream subreddits, saying they hold the favour of the admins are therefore able to influence and change the ecosystem as they please whilst disregarding repercussions for smaller niche subreddits, “who are acting within their own interests as this, like. It feels like, they could be influencing us in ways that we can't even see or anything about. And it's kind of spooky to think about it. I mean, yeah. I'm not sure if you're whereas like my cozy, like subreddit gets one it's like people are scheming to change things or getting rid of a certain people , like user base that they didn't

like, and they've been given the tools to do it. That's the thing they've been given the tools to do it. Um, like I said, it's an internal moderation. Like these moderators got to like watch out for themselves".

The theme thus depicts that smaller niche, non-mainstream subreddits have closer ties to users, especially those that have an intrinsic understanding of the culture, whilst larger subreddits have a larger focus on the profit and political structure of Reddit. Furthermore the disposition explained by interviewee 8, describes that these actors are comfortable putting themselves at odds, with an "us versus them" mentality, with no worries or care because they hold the power in the relationship. This thus outlines the structure as moderators giving the maximum allowed freedom to the user of the space and tools that are given to them by the admins.

Controversial content indicator – 4.4

The theme of controversial content indicator refers to the observable aspects of content that can make it controversial. This theme helps define the data that has a focus on disputes between values and beliefs that eventually produces a controversial topic, whether directly or indirectly. As the legal framework exists and struggles to define what this theme encompasses, moderators that derive from communities and deal with this content on the regular and can differentiate what content is meaningfully controversial and how it should be approached.

From the interviews, it is clear that controversial content is a deeply complex issue that exists on several levels, and can be approached from several different angles. The conversation in the interviews pivoted from what is controversial, to how to approach the controversial. The understanding is that no one can choose what is controversial, because anything can essentially be controversial. A global platform allows anyone, who lives in a different background, in a different culture with different experiences to find some form of content to be controversial.

The only unanimous agreement, with the exception of interviewee 2, was on illegal content, which the participants would separate it from controversial, as illegal had clearer effects, citing when it is obviously hateful, traumatic or dangerous, triggered by extreme elements such as nudity or gore, some drawing the line even further when it was obvious it would spill over into the real world. For example Interviewee 1 stated "Um, But like, right, what I would say is when things that eventually could lead into real life ramifications, like calls for violence, targeted harassment. Um, and I think probably because of the type of people that could congregate and illict reactions such as like domestic terrorism, pref- preferably like, you know, racist or like extremist material., I'd say that's probably the, the line I think," In those

instances moderators would remove the content. Controversial content on the other hand does not necessarily need to contain extreme elements instead is more defined by the type of reaction it generates.

As interviewee 10 explained in 2 instances “Uh, yeah, I think it's unavoidable,. Um, because like I said, why I feel controversial, content is stuff that gets extreme reaction, like emotional reactions.”, and “Um, for me, it's, probably something that gets its content, they get lots of emotional reactions, let's say, uh, sometimes which leads to opportunities for problems such as like fake news to happen.”, and interviewee 6 mentions, “Um, I'd say controversial content is something that could potentially cause a stir or negative reaction. Um, I'm not se- yeah on like a negative reaction. It could be anything, I guess it could be like set comment or it could be a post entirely, or it could be something else. Like, yeah. It could be someone like a moderator has done a thing in private that is bad or not. It could be considered controversial content”

The stance interviewee 2 took, was that in some instances extreme content should be permitted, following a radical media literacy perspective. As quoted from the interview, “I mean listen, I’m a jew and I've been at right spaces politically. So honestly, I think just as long as it's not illegal, I almost actually even possibly argue that even like ISIS beheading should be allowed simply because to show just the brutality of these people, you know, and just how important is and why we're fighting against them, you know, like, because what ends up happening even in terms of hate speech or whatever, you'd want to classify it as, like, I get a bunch of people who I interact with they don't like Jews, you know. Even to the extent that, Yeah, this other guy who's Jewish and they up each other up about how we have invented and the space, laser and stuff like that, and just joke around about it. It's not that deep at the end, you know, they're allowed to have whatever opinions and be immediate. Yeah. That's just my opinion.”

In essence, when a user posts, the intention of a post may be harmless but may cross another user on the site who deems it offensive. It was described by Interviewee 5, as “controversial content doesn’t exist, instead it appears”. Thus only strategies are implemented to regulate and contain it when it surfaces. Tools such as NSFW (Not Safe for Work) tag gives the user autonomy and less of a responsibility to the moderator to define controversial content. If the user decides to open the tag, they have exposed themselves to the material, whether it offends them or not. On the moderators side, it becomes more of a personal issue in which the ethical and philosophical values of the moderator comes in deciding the validity of its

controversy. Interviewee 5 goes as far to state “It is a moral duty that requires a perspective on empathy and sympathy because it is less about the content and more about feeling insulted.”

Even if the interviews suggest controversial content is unable to be defined, it can be identified in volume according to subjects and genres as certain topics generate more emotional and negative reactions. For example interviewee 1 indicates this by saying “And obviously this is just a game, but if you were like a political subreddit, the subreddit that people will take it a lot more seriously to the things you moderate.”, implying explicit moderation is more likely to occur as a result from the seriousness of topic with users being more emotionally invested and opinionated because of it.

Another interview indicated how content can be controversial by nature, based on interviewee 7, that is moderator of r/ChrisChanSonichu. This is because the topic of the subreddit revolves around an individual who is highly controversial, and therefore any content related to it is guilty by association regardless of context. Thus as a result, every single post was tagged as NSFW on the subreddit even if no typical controversial elements are present, neither in text nor in image. Since the subreddit rules require posting to be related to the controversial individual, the connection is enough to make the content posted controversial by nature. The exchange that occurred during the interview confirms this, Interviewee, “controversial content is our content. It's is. Yeah.”, Interviewer, “…you're guilty by association because of the person and these things that anything you post on the subreddit itself would indicate that it's directly related to this controversial person. Hence it is controversial content, right?” Interviewee, “yes”.

The theme ultimately demonstrates that controversial content has several angels, approaches and levels that it needs to be considered from. As it currently stands, it can not be clearly defined, and only properly dealt with on a reactionary basis.

Human-Bot Collaboration moderation - 4.5

This theme includes part of the data that focuses on the auto-moderator of Reddit and its effectiveness in moderating alongside human moderators. This ranges from the hands-on tools and abilities to its agency within the subreddit.

The interviews revealed that human-bot collaboration on moderation, remains a majority human-driven activity, with the desire remaining so. Interviewee 2 stating “Yeah, I think they should definitely take a secondary role of alteration. I think it should be primarily a, uh, human driven activity.”. Further reasoning on his behalf was that the increased use of automated tools, instead of aiding would instead do the opposite and present itself as an

obstacle, “. I generally have kind of a negative opinion of the use of those tools. I find that more often than not, they cause problems when they're using a very limited sense. I”. Interviewee 5, was also uncomfortable with that notion and voiced “ugh, I dunno because it's like automated systems. They're not perfect either. They can always make mistakes and a lot of the time, no one's held accountable.”.In this case, subreddit culture and topic had a much smaller impact too as the role of bots and the auto moderator is primarily used as a filtering and notification tool, regardless of content. The answers of the participants remained consistent between every single interview, and answers regarding this theme remained relatively short, revolving around moderator knowledge of automated tools and the philosophy that automated tools lack the nuance to effectively moderate content.

The first consistency in answers was how bots and automated tools were limited because most moderators lacked the knowledge to use them efficiently, often being left to a single moderator in the network knowing how to operate it. As a result, that one moderator is delegated tasks when alterations in the functions of the tools are desired. This is evident for interviewee 4, stating “Yeah, I came up with the automod, how it was set up. Um, I have some knowledge of it just because I set it up before. Yeah. Um, you know, I'm not very knowledgable about that.. I mostly use it just as like a way to like link to other resources, stuff like that.” And another example where Interviewee 6, was asked about moderator collaboration with automated tools responded, “Yeah. I've been a little moderatoring for your price though. Don't really know how it works. Um, I get stuck on something to kind of leave it, pretend it doesn't exist., but now, yeah, it's, it's pretty true what they say about like most moderators don't even know what hell it does.”.

The second consistency in the theme was how subreddit size effected the likeliness to using the automated functions. Moderators speculated that larger subreddits with higher traffic, users and active visitors were compelled to use bots and automated tools as volume would be too much to handle, between the ratio of influx of content and human moderation. All 10 moderators responded similarly, that this was likely the case, however could not confirm if this was true and would only speculate based off of their own personal experience in moderation. Interviewee 2, for example stating, "So I just have to speculate here. What I would suggest is for those larger subreddits, with the constant churning content that they use a combination of, uh, automated tools, bots,...". Interviewee 9 stating, “It seems like an inevitability that people would have to steer towards having automated moderation. They could have our own moderation or on the end, like at 400 K or something like that, where like other subreddits, they're like, well, over a million yeah. So almost like a massive like a city essentially.”

In essence, the moderators believe bots lack to nuance to moderate content effectively, alongside the lack of knowledge, or motivation to learn about the tools delegating it to the most knowledgeable moderator in their network. Only under circumstances such as rapid growth and influx of content would motivate them to learn and implement the bots.

Ethical and philosophical framework – 4.6

This theme of the ethical and philosophical framework makes sense of the data that refers to the personal beliefs and values that the moderator holds themselves. This includes how the moderator distinguishes their moderation style in relation to the restrictions that are placed on them and how they apply or modify their values to fall in line with the required limitations, whether it be a more implicit or explicit approach.

Earlier chapters have already indicated how the ethical and philosophical framework of the moderator plays a major part in determining moderation style and the view on the importance of nuance. In this case, the theme itself was more focused on particularly how the moderators distinguished their beliefs that may have effected moderation choices and beliefs in the space. This can be commentary in the form of lived experiences outside of moderation and how it affects their current thought pattern, or how their beliefs came about whether it is on moderation or not. Moderators of the watchdog subreddit (WRD) were more vocal and motivated to share their beliefs, e.g. Interviewee 4, self-described “Free speech absolutist” as explained before.

And another example by interviewee 8, with a reply on their stance and motivation for moderating controversial content “Um, no. Yeah. And the reason I bring this up is because I think that, uh, at least regarding the stuff I'd have to deal with on the discord server, I think that, uh, this type of rhetoric, and I guess epistemology or, you know, street epistemology. Um, it, I guess it's just one of those things that I always kind of like understanding, you know, how do we do. People view the world because it's a pretty big world. There's a lot of new people, you know what, I'm stuck with everyone and I want them to have, but at the same time, you know, learning a different way to interpret the world can be pretty fun.”

The remaining respondents did not add as much emphasis on their ethical and framework towards their moderation and mainly trusted a process of discussion among their network for moderation of content with a primarily objective outlook.

Reddit Ideology Integrity – 4.7

This theme is makes sense of all the data that would portray Reddit and its values as it is displayed from the outside compared to inside operations, according to the moderators. This includes their lived experiences with the website and the culture and ideology it is believed to have according to its supposed conduct and practices which includes the effects of admins, moderators and users that traverse the site.

Overlap of the theme is indicated as earlier chapters reveal how Reddit portrays itself, versus how it actually operates according to moderators of the community. However instead of moderators implying its ideology through cases, this theme refers more to the direct opinions. For example, Interviewee 7 both understands and critiques the Reddit insincerity, but still appraises the space, for its ability to allow most types of niche content that better connects its nuanced culture to the mainstream "I can't imagine online life without place to discuss certain things, whether or not they're niche or not. It is kind of hard to imagine. Um, yeah, in my opinion, yeah. I can reddit as well, like even, yeah. our subreddit is small.".

Other respondents focus more on enforcing its ideology for profit, in which interviewee 8 exclaims a case about Reddit not enforcing any political agenda as it is stereotypically perceived as a politically liberal space, stating "I should say in there. Kotakuinaction is usually right-leaning um, I will say, I don't know why they put it back, but I wouldn't be surprised if it is a money maker." (in reference to how Reddit admins re-added a right wing subreddit because it was profitable).

This response directly contradicted the answer of interviewee 4, who stated the ideology was not about money but enforcing a political ideology, by stating the case that right-wing subreddit T_D was banned despite being one of the most profitable subreddits. "I think it's less money and more the power that they got the power being like this sort of massive head almost."

Reddit Moderator Culture – 4.8

The theme of Reddit moderator culture focuses on the data depicting the customs, ideology and principles moderators find themselves on. This is a niche in between the personal ethical framework of the individual and the structure in place provided for them by the administrators. Reddit moderator culture depicts another angle in how interactions between teams of moderators turn and shape their moderations style and how it is affecting the space when done. It also highlights data that focuses on the different moderator cultures that exist and each one affects the perception of online spaces.

The first set of evidence after analysis of the data revealed that moderation culture on Reddit is very fractured. This is as a result of all 10 moderators separating themselves from an apparent mainstream Reddit moderator stereotype. As interviewee 1 points out “Um, I think there was a reputation among Reddit for that type of thing, but like I said, uh, subreddits are kind of niche and, small... ...maybe we'd have one person that was a little bit too hands-on but a lot of the time, that's not really something to be noted about compared to other subreddits where people are really curbing discussion with things that we don't agree with.” And Interviewee 2 stating “I think another part of it, uh, I think a lot of moderators kind of the power gets to their head as silly as that sounds… …Knowing that you could be part of that. I think that gives them a sense of power and it's just too tempting to use that. And if they have a particular agenda, they like particular viewpoint, they want to promote, um, they just immediately jumped to that ban hammer or, uh, you know, promoting sticking comments.” Overall, it indicates that moderators of larger mainstream subreddits are more prone to feeling a sense of importance and power as their decisions effectively influence more peoples.

Next, moderator culture revealed the concerns around mental health of individuals moderating. Evidence provided in the chapters above show how some moderators deal with content that include elements of gore, nudity and violence that is not appropriate for the average user. This can be exacerbated by the time input of some moderators that spend up 10 hours facing such content. Interviewee 3, went on to voice “It does, I mean, I've seen, I mean, like there's some vice documentary, so of actual Reddit moderators right. Is that live in pretty bad conditions because obviously they're dedicating their entire day to moderating and you know, the house is dirty they haven't showered I mean, yeah. It's sure it's pretty sorry. Um, yeah, it's pretty difficult and it definitely gives enough. Uh, yes. As a scale.”.

Interviewee 8, exclaimed how he is fine dealing with extreme content because of his condition of Aphantasia, by which he can’t visualize images in his head properly. “It's one of the things that's like, you can't really say anything worse to me about. That I haven't already heard. Um, I will say that like nobody's ever tried to dox me or try to actually threaten me. So in the chance that those happen, that might take more of a mental too.”, therefore implying moderators are harassed on scale by doxing and verbal threats. Interviewees imply that the more they moderate, the more they become desensitized by content.

Insincerity – 4.9

The Theme of Insincerity focuses on corruption, double standards and insincere attitudes admins or other moderators display on the website according to the moderators. It highlights the severity of actions taken to by those actors and the resulting consequences of those cases, revealing the deeper flaws of the internet structure as a whole. Insincerity was a consistent theme among all moderators, that not only watchdogs subreddit moderators identified these issues.

This theme had the most distressing answers, as interviewees explained several cases implying how little official institutions and admins affect the space to promote for a safe discussion. Interviewee 4 even went as far as to accuse another subreddit and their moderators of posting illegal content to other subreddits they dislike to get it taken down. “you know, um, you have stuff like against hate subreddits, which has a long documented history. Of using child pornography to shut down other boards, other subreddits stuff like that. And yet they're still allowed up, even though they violate, you know, anti-brigading rules that are on the platform.” The insincerity comes the inaction of Reddit admins, which have not investigated the users responsible for engaging in these practices. Even more concerning to interviewee 4 was how no legal consequences occurred for users who had been identified to post such content, as entire communities were removed on the basis of bad actors. It highlights the minimal impact the current authoritative institutions have over the web. According to him It does not reach deep enough and only impacts on a surface level.

The data further depicts that Reddit holds a double standard and has a hypocritical attitude towards specific subreddits. 3 interviewees, 1, 2 and 4, are convinced Reddit is enforcing a political agenda, because it holds subreddits to different standards. An earlier chapter already discussed the banning of ‘T_D’ whilst keeping the subreddit ‘BernieBros’ who committed the same violation but were allowed to remain on the website. These were discussed in various instances as for example interviewee 1, explained “It depends on the topic as well. Cause I know like certain subjects, like female dating strategies, um, a lot of the time. It's like primarily all the time, like complaining about men and like disparaging them be like, is that sexism? Is that sort of discrimination?”, and a further citation by the moderator had shown other examples of subreddits, such as ‘againstmensrights’ and ‘fragilewhiteredditor’ existing to discriminate and criticize men and white people, but when searching for ‘againstwomensrights’ and ‘fragileblackredditor’, they were banned and/or removed.

This stance cannot be entirely accepted as truth to the situation, as Reddit also evidently is known for lacking transparency in their actions. Interviewee 4 admits to themselves “it's a

combination. It's the discrepancy of moderation and it's the lack of transparency." And Interviewee 5 elaborates on this with "to be honest, I have no idea what the hell is going on, which probably speaks a lot to like their transparency."

Other external unknown factors could have contributed to the ban of mentioned removed subreddits. However, as of yet, no further explanation or elaboration on these events have been afforded. Another argument made on the conduct occurred from interviewee 8, who in the same network as interviewee 2 and 4, contradicted their speculations and explained that it is within their legal rights as private companies to conduct such actions and hold a political stance by saying "Whether intentionally or not conflate the ways the different rules work. So that way they can continue their Neo Tillman. I picked them what complex. And I say that because it's one of those things that's like, you know, after years of hearing conservatives say things like, you know, you're not entitled to anything they'll suddenly go full 180 and be like, I'm entitled to a Twitter account. I'm like, what the hell are you on?"

The theme of insincerity proves that the structure of Reddit holds flaws that has dangerous outcomes if users are permitted to be allowed acts such as post illegal content, with no harsher repercussions.

Social commentary – 4.10

The Theme of Social commentary is the data that has the moderator provide commentary on how the current affairs of the system of the online public sphere affect the outside real world and vice versa. This can often be commentary on how the capitalist and/or political exploits affect the spaces or systems; or where the attention and resources are focused to whilst overlooking other vital elements that further hamper the progress and improvement of the online spaces.

The results of this theme shows that moderators believe that regulating online content through authoritative and legal institutions is futile. The first reason is the belief that the generations in charge of creating these laws are ill equipped to fully grasp the volume and circulation of online content and its boundaries. Interviewee 6 for example states, "The old generation has a misunderstanding we don't even need like that paradox Plaza to, um, to discuss everything because that understanding and that culture will already be ingrained within, you know, society" and interviewee 1 says, "Because obviously the amount of people like users are inevitably re increasing. We have a younger generation that are familiar with digital technology" (in response to if Reddit will suffer consequences to their userbase from their inaction on insincere actions).

Another result of observing the data in this theme is the critique of capitalism. As structures of the website are set around that foundation, according to 7 of the respondents, with 1 opposing this idea, assuming structures are more politically oriented. Other themes have depicted the case studies relating to this notion. In direct commentary on ideological perspectives, for example interviewee 8 voiced “You know, they want, it's basically like, you know, these things, they are businesses. They are trying to make money. And while every decision they make is not going to be the right decision or even the most profitable one, um, their main goal is to appease their stockholders or the shareholders or whatever the official term is. Um, or w or even just the owners, like it's a privately traded company.” And interviewee 5 citing, how their inaction could be a choice in itself, as “Yeah. I mean, everyone controversies happened probably proper profitable to Reddit.”

Interviewee 7 stated a tactic that would be more effective in bringing change into these platforms, as an informal procedure to incentivize change. Understanding flaws in the system may not necessarily mean changing the system but adapting tactics that apply to the system, “Action would be making a public deal off of, cause I think a corporation would actually kind of care because there's the people that using the platform and actually making the money.” In this instance the very criticisms these moderators voice on the current systems also warn that those which make use of those can also suffer the same consequences they advocate for in the system. “So you're removing this because, and at the end of the day, it comes down to, they think that they're always going to be empowered power. You never see how the tools that they're currently using, can be used against them..” (interviewee 4, in response to a question about preventing disinformation).

The theme of Social Commentary encompasses the wider debate how ideology structures the online public spaces, and how it is being exploited in such a manner. Moderators feel at a structural impasse as there is too much corruption to effectively change the ecosystem for the better.

Chapter 5

Conclusion – 5.0

In conclusion, the theory regarding online moderation depicts motivations as authoritative and institutional influences combined with personal values of the moderator for the Reddit ecosystem. Yet as results reveal that the authoritative structures are inadequate to respond to illegal content, enforcement of healthy online spaces and real-life harassment. Along with the structure of content spaces within Reddit being too corrupt and ineffective to provide meaningful change. Several instances of illegal content that is posted which remains unpunished by Reddit or other authorities of users responsible. With the result for communities to suffer under bad actors with ill intentions, whether it is for political or other intentions. Moderators feel exposed and opposed to by admins rather than supported in these spaces, which is even reflected by the manner in which they approach outside sources with information, such as this research project, resulting in the use of voice scramblers. It was inferred that only true change may occur if there was a shift in dialogue to nuance.

Limitations – 5.1

Limitations in the research regarded mainly to sampling size and initial contact to the participants. As explained, due to the platform being dominated by western consumers, all moderators that participated were either descendent from north America or western Europe. Furthermore the mentioned weariness faced when contacting this group left many criteria unidentified, such as age, sex and name. Furthermore several obstacles occurred as it was unexpected that moderators had the function to block direct messaging in Reddit chat and Direct Messaging on the platform, resulting in only 1 function to reach them through the 'Message all the moderators' button. Even this function does not reliably reach the moderators as it was admitted by several that the mod mail is not reviewed often resulting in rare replies. Posting about it in the respective subreddit will result in distrust as communicating in such a manner breaks the rules. Hence when successful contact is achieved, snowball sampling is the most viable option, but then increases the likeliness of biased results.

References

Achimescu, V., & Chachev, P. (2020). Raising the Flag: Monitoring User Perceived Disinformation on Reddit. Information, 12(1), 4. https://doi.org/10.3390/info12010004

Almerekhi, H., Jansen, S., & Kwak, c. (2020). Investigating Toxicity Across Multiple Reddit Communities, Users, and Moderators. Companion Proceedings Of The Web Conference 2020. https://doi.org/10.1145/3366424.3382091

Anna Gibson. 2019. Free Speech and Safe Spaces: How Moderation Policies Shape Online Discussion Spaces. Social Media + Society 5, 1 (2019), 2056305119832588.

Barr, R. (2019). Growing Up in the Digital Age: Early Learning and Family Media Ecology. Current Directions In Psychological Science, 28(4), 341-346. https://doi.org/10.1177/0963721419838245

Breakstone, J., McGrew, S., Smith, M., Ortega, T., & Wineburg, S. (2018). Why we need a new approach to teaching digital literacy. Phi Delta Kappan, 99(6), 27-32. https://doi.org/10.1177/0031721718762419

Çömlekçi, F., Güney, S. (2014). An Alternative Media Experience: LiveLeak. In: Marcus, A. (eds) Design, User Experience, and Usability. User Experience Design for Diverse Interaction Platforms and Environments. DUXU 2014. Lecture Notes in Computer Science, vol 8518. Springer, Cham. https://doi.org/10.1007/978-3-319-07626-3_6

Conner, M., Perugini, M., O'Gorman, R., Ayres, K., & Prestwich, A. (2007). Relations Between Implicit and Explicit Measures of Attitudes and Measures of Behavior: Evidence of Moderation by Individual Difference Variables. *Personality And Social Psychology Bulletin*, *33*(12), 1727-1740. https://doi.org/10.1177/0146167207309194

Clark, D. M., & Wells, A. (1995). A cognitive model. Social phobia: Diagnosis, assessment, and treatment, 69, 1025.

DE STREEL, A., Defreyne, E., Jacquemin, H., Ledger, M., & Michel, A. (2020). Online Platforms' Moderation of Illegal Content Online: Law, Practices and Options for Reform. European Parliament. 1(1), 25-39

N. Facione, P. Facione (2020) Measure what matters: Critical thinking skills & mindset. (n.d.)Insight Assessment

Facione, P. A., Gittens, C. A., & Facione, N. C. (2016). Cultivating a critical thinking mindset. Academia. Edu. Weekly Digest, 28.

Fuchs, C. (2013). *Social media and capitalism* (pp.25-61). Nordicom.

Geiß, S., Magin, M., Jürgens, P., & Stark, B. (2021). Loopholes in the Echo Chambers: How the Echo Chamber Metaphor Oversimplifies the Effects of Information Gateways on Opinion Expression. *Digital Journalism*, *9*(5), 660-686. https://doi.org/10.1080/21670811.2021.1873811

Glenski, M., & Weninger, T. (2017). Predicting User-Interactions on Reddit. *Proceedings Of The 2017 IEEE/ACM International Conference On Advances In Social Networks Analysis And Mining 2017*. https://doi.org/10.1145/3110025.3120993

Hill, K. (2022). *Reddit Co-Founder Alexis Ohanian's Rosy Outlook On The Future of Politics*. Forbes. Retrieved 23 June 2022, from https://www.forbes.com/sites/kashmirhill/2012/02/02/reddit-co-founder-alexis-ohanians- rosy-outlook-on-the-future-of-politics/.

Illegal content on online platforms. Shaping Europe's digital future. (2022). Retrieved 23 June 2022, from https://digital-strategy.ec.europa.eu/en/policies/illegal-content-online-platforms.

Jasser, J., Garibay, I., Scheinert, S., & Mantzaris, A. (2021). Controversial information spreads faster and further than non-controversial information in Reddit. Journal Of

Computational Social Science, 5(1), 111-122. https://doi.org/10.1007/s42001-021-00121-z

Jhaver, S., Appling, D., Gilbert, E., & Bruckman, A. (2019). "Did You Suspect the Post Would be Removed?". *Proceedings Of The ACM On Human-Computer Interaction*, *3*(CSCW), 1-33. https://doi.org/10.1145/3359294

Jhaver, S., Ghoshal, S., Bruckman, A., & Gilbert, E. (2018). Online Harassment and Content Moderation. *ACM Transactions On Computer-Human Interaction*, *25*(2), 1-33. https://doi.org/10.1145/3185593

Jo Lander. 2015. Building community in online discussion: A case study of moderator strategies. Linguistics and Education 29 (2015), 107 – 120.

Kahneman, D. (2011). Thinking, fast and slow. Macmillan.

Kahne, J., & Bowyer, B. (2018). The Political Significance of Social Media Activity and Social Networks. *Political Communication*, *35*(3), 470-493. https://doi.org/10.1080/10584609.2018.1426662

Kearns, I., J. Bend, and B. Stern. 2002. E-participation in Local Government. London: Institute for Public Policy Research.

Kendall, L. (2008). The conduct of qualitative interviews. *Handbook of research on new literacies*, 133-149.

Langvardt, K. (2017). Regulating Online Content Moderation. SSRN Electronic Journal, 1(1), 1335. https://doi.org/10.2139/ssrn.3024739

Loader, B., & Mercea, D. (2011). NETWORKING DEMOCRACY?. Information, Communication &Amp; Society, 14(6), 757-769. https://doi.org/10.1080/1369118x.2011.592648

Manning, N., Penfold-Mounce, R., Loader, B., Vromen, A., & Xenos, M. (2016). Politicians, celebrities and social media: a case of informalisation?. Journal Of Youth Studies, 20(2), 127-144. https://doi.org/10.1080/13676261.2016.1206867

U. Matzat and G. Rooks. 2014. Styles of moderation in online health and support communities: An experimental comparison of their acceptance and effectiveness. Computers in Human Behavior 36 (2014), 65 – 75.

Medvedev, A., Lambiotte, R., & Delvenne, J. (2019). The Anatomy of Reddit: An Overview of Academic Research. *Dynamics On And Of Complex Networks III*, 183-204. https://doi.org/10.1007/978-3-030-14683-2_9

Meneses, L. (2021). Thinking critically through controversial issues on digital media: Dispositions and key criteria for content evaluation. Thinking Skills And Creativity, 42, 100927. https://doi.org/10.1016/j.tsc.2021.100927

Mills, R. (2011). Researching Social News–Is reddit. com a mouthpiece for the 'Hive Mind', or a Collective Intelligence approach to Information Overload?.

Mills, R. A. (2018). Pop-up political advocacy communities on reddit. com: SandersForPresident and The Donald. Ai & Society, 33(1), 39-54.

Ovadia, S. (2015). More Than Just Cat Pictures: Reddit as a Curated News Source. Behavioral &Amp; Social Sciences Librarian, 34(1), 37-40. https://doi.org/10.1080/01639269.2015.996491

Pierce, L. (2022). Dave Chappelle's Sticks and Stones as Black Radical Tragic comedy. *Text And Performance Quarterly*, *42*(2), 126-143. https://doi.org/10.1080/10462937.2022.2036803

Potter, M. (2021). Bad actors never sleep: content manipulation on Reddit. *Continuum*, *35*(5), 706-718. https://doi.org/10.1080/10304312.2021.1983254

Reddit.com. (2022). Retrieved 23 June 2022, from https://www.reddit.com/.

Reddithelp.com (2022). Retrieved 23 June 2022, from https://mods.reddithelp.com/hc/en-us.

Schäfer, M., Füchslin, T., Metag, J., Kristiansen, S., & Rauchfleisch, A. (2018). The different audiences of science communication: A segmentation analysis of the Swiss population's perceptions of science and their information and media use patterns. *Public Understanding Of Science*, *27*(7), 836-856. https://doi.org/10.1177/0963662517752886

Sevignani, S. (2015). *Privacy and Capitalism in the Age of Social Media*. (pp.12-32) https://doi.org/10.4324/9781315674841

Stolworthy, J. (2022). Retrieved 20 June 2022, from https://www.independent.co.uk/arts-entertainment/tv/news/rowan-atkinson-netflix-cancel-culture-b2104759.html.

Van Dijck, J., Poell, T., & De Waal, M. (2018). *The Platform Society* (pp. 46-75). Oxford University Press USA - OSO.

Williams, J. (2019). The use of online social networking sites to nurture and cultivate bonding social capital: A systematic review of the literature from 1997 to 2018. *New Media &Amp; Society*, *21*(11-12), 2710-2729. https://doi.org/10.1177/1461444819858749

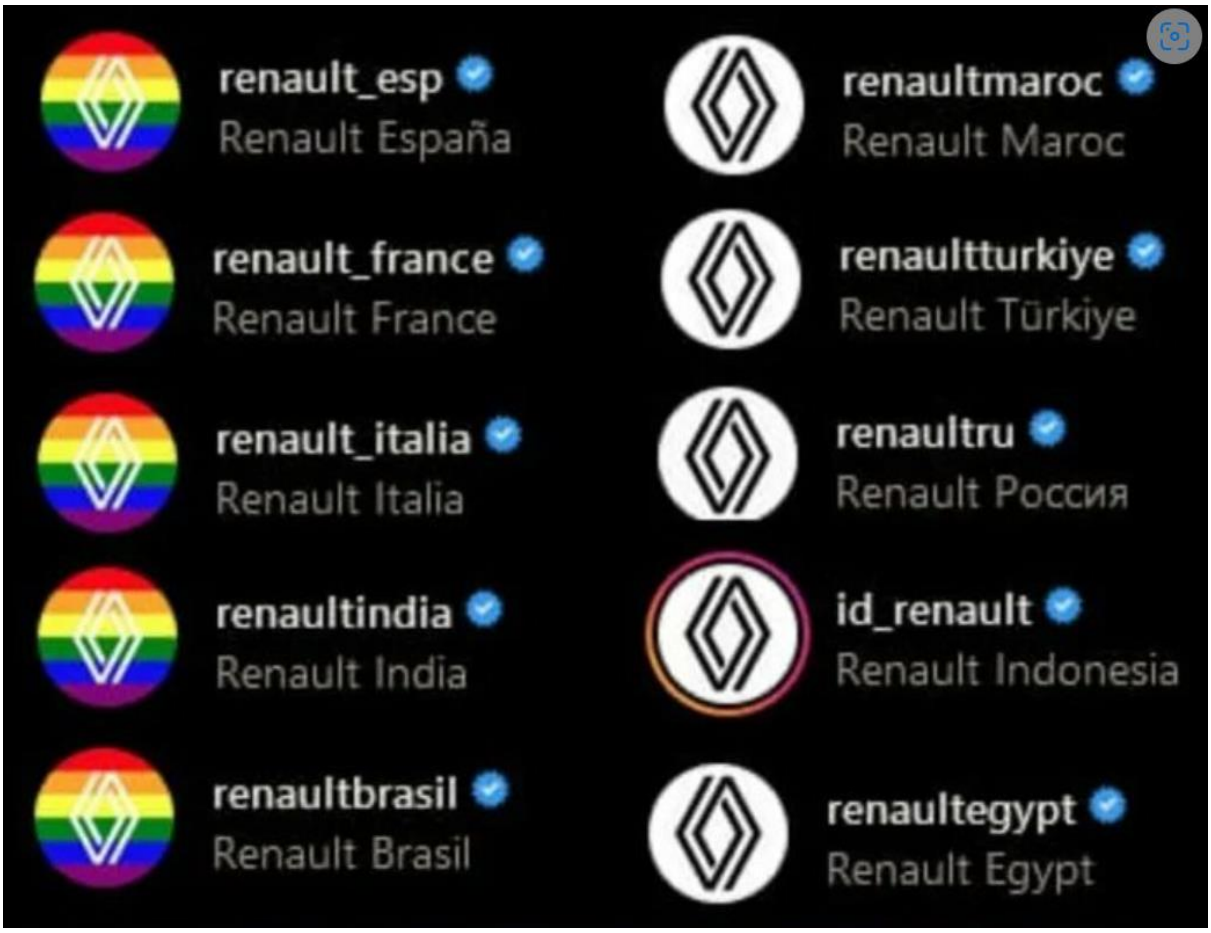
Weinberg, H. (2018). *The paradox of internet groups* (3rd ed., p. 19).

Wright, S. (2009). The role of the moderator: Problems and possibilities for government-run online discussion forums. Online deliberation: Design, research, and practice, 233- 242.

Appendix

Appendix A.

Renault Logos Example



Appendix B.

Participant Information

Interviewee #	**Moderator Handle**	**Genre**	**Subreddit name**	**Sex**
1	X	Gaming	r/Mordekaisermains	Male
2	u/yeahyeahhright	Watchdog	r/WatchRedditDie	X
3	X	Comedy	r/GreenText	Male
4	u/Fe3rless	Watchdog	r/WatchRedditDie	Male
5	X	Gaming	r/Warhammer40k	Male
6	X	Gaming	r/Sterllaris	X
7	X	Controversial	r/ChrisChanSonichu	Male
8	X	Watchdog	r/WatchRedditDie	Male
9	X	Entertainment	r/AOT	X
10	X	Gaming	r/GhostOfTsushima	Male

Appendix C.

Final Coding Results

Figure C 1: Coding Scheme

# Themes	Relevant Codes	
1. Moderation Style	Moderation Style	colour
	Discussion incentive	colour
	Legal	colour
2. Nuance	Nuance	colour
	Rules	colour
	Subreddit Culture	colour
	Subreddit Size	colour
3. Interpersonal Relationships	Interpersonal Relationships	colour
	Third Party Relationships	colour
4. Controversial Content	Controversial Content	colour
	Consequences	colour
5. Human-Bot Collaboration Moderation	Human-Bot Collaboration Moderation	colour
	AI and Algorithms	colour
6. Ethical and Philosophical Framework	Ethical and Philosophical Framework	colour
	Structural change	colour
7. Reddit Ideology	Reddit Ideology	colour
8. Reddit Moderator Culture	Reddit Moderator Culture	colour
	Recruitment	colour
	Motivation	colour
9. Insincerity	Insincerity	colour
10. Social Commentary	Social Commentary	colour

Figure C 2: Counted Codes and Results

Themes	Codes	i.1	i.2	i.3	i.4	i.5	i.6	i.7	i.8	i.9	i10	Total	Total Themes

Moderation Style	Moderation Style	10	2	7	8	11	13	8	14	9	12	94	105
	Discussion incentive	1	0	0	0	0	1	0	0	0	3	5	
	Legal	0	2	1	0	0	0	0	3	0	0	6	
Nuance	Nuance	26	6	21	2	11	15	11	4	15	16	127	169
	Rules	1	1	5	1	0	1	0	0	1	1	11	
	Subreddit culture	6	3	4	0	4	3	3	0	1	0	24	
	Subreddit Size	0	0	3	0	2	1	0	0	1	0	7	
Interpersonal Relationships	Interpersonal Relationships	11	5	7	7	8	13	1	7	9	9	77	81
	Third-Party Relationships	0	1	1	1	1	0	0	0	0	0	4	
Controversial Content	Controversial Content	6	6	9	3	5	8	13	6	6	8	70	85
	Consequences	0	0	2	4	1	3	2	0	0	3	15	
Human-Bot Collaboration Moderation	Human-Bot Collaboration Moderation	7	9	5	7	4	5	3	4	1	6	51	53
	AI and Algorithms	0	0	0	0	0	0	1	1	0	0	2	
Ethical and Philosophical Framework	Ethical and Philosophical Framework	10	16	6	12	3	5	7	17	8	8	92	108
	Structural change	4	11	1	0	0	0	0	0	0	0	16	
Reddit Ideology	Reddit Ideology	13	22	6	6	8	9	4	10	7	16	101	101

Reddit Moderator Culture	Reddit Moderator Culture	6	15	15	5	5	2	8	7	7	8	78	105
	Recruitment	6	1	1	1	2	0	0	2	1	1	15	
	Motivation	0	1	1	1	2	2	0	2	3	0	12	
Insincerity	Insincerity	0	4	5	18	5	2	4	3	5	16	62	62
Social Commentary	Social Commentary	0	1	5	6	1	2	7	37	10	0	69	69

Appendix D.

Coding Tree

Figure 1 D.

Selective	Axial	Open
Moderation Style	Discussion incentive	Facitilitate productive conversation
	Explcit moderation	Prevent disruption of ordinary
	Implicit moderation	Make mega thread for discussion
	Moderation Incentive	Mix is in style of moderation
		Implicit Moderation: How can it be to the benefit to users, benefit of the doubt
		Implicit moderation encouraged moderation practice by Reddit
		Role of the moderator
		Rules used to stay on task
		Explicit moderation occurs when users continuously violate rules
		Explicit moderation can harm foundation of the subreddit
		Personal decision, moderation is a personal decision
		Priority list determines who or what is moderated first
		Explicit moderation occurs according to guidelines of service and community
		Explicit moderation is deemed counterproductive
		Moderate based on how admins react
		Rules: Clear, Straightforward, Consensus
		Topic defines subreddit culture
		Explicit moderation is generally used minimally and is only enforced according to rules
		Prioity to enforce rules but can sometimes let things slide/Bend rules
		Dependent on moderating according to external but relevant actors I.e. Paradoxplaza
		Focus on having a good time
		Nature of content can also require a predominantly explicit moderation style
		Defaulting content to a new standard
		Innovative moderating strategies/playing the system
		Moderation can highly be based off of conduct of user
		Rules are set to make sure things won't become complicated
		Real staff of actual game (Stellaris/Paradox) involvement, explicit

Figure 2 D.

Selective	Axial	Open
Nuance	Media literacy	national governmnt restrictions
	Subreddit Culture	totalian countries will want content removed
	Rule enforcement	Apply restrictions to hate speech about holocaust
	Misinterpretation	Comments or discussions deemed/mistaken as hate speech
	User perception	Blanket statements and ad hominem vs. nuance
		"Immigrants are ruining the country, can actually be about political policies and societal problems"
		User's don't want 3rd party interference
		Hard to explain to someone who is not part of internet culture
		Dependent on Subreddit culture either to be implicit or explicit
		Lenient because of content they deal with
		Not noticed compared to other subreddits
		Sub cultures tend to bring in a "certain type of person" hence politics is controversial
		"If you know, you know" - on tackling internet hate speech strategies
		"Moderation should remain human-driven"
		Always need a human moderator on hand to provide nuance for given content, I.e. humour, questions and controvers
		Attitude severely impacts how moderation is done and viewed
		Subreddit culture, topic niche makes complaints about moderation on such a sub unnoticed because they are regulars
		Societal pressure on niche topics people have no understanding of, unfair rules
		Nuance is determined by added actors
		Sub culture, based on certain actors, "Just how we are", "lack content"
		Diverse moderation team to get all perspectives
		Spaces are required to understand topics ona deeper more critical level
		Trolls are prevalent in online forums especially in controversial content spaces
		Focus on having a good time for sub, in relation to more serious variant (Stellaris/Paradox)

Figure 3 D.

Selective	Axial	Open
Interpersonal relationships	User-Moderator relationship	Interact regularly with user-base
	Reddit Admins-Moderator relatioship	Regulars in a sub, "flawless trackrecord", no bad user history
	Mod-Mod relationship	Familiar with relationship to users
	Third-party relationships	those voicing issues with subreddit, rules or moderations are usually non-regulars (niche subreddit)
		Subreddit needed moderators, colunteering to assist in reviewing
		There is either an understanding or you're their enemy
		Free hand for moderation = corrupt power
		Big subreddit moderators tend to face more bad actors which leads to explicit moderation
		Mods play a backgorund role, not noticed much
		Let users mark posts themselves
		How users use their autonomy determines their access to the subreddit (ostracization)
		not all mods agree on moderation method, discourse, democratic
		Nepotism, Passionate, Ideological
		Subreddit held to different standard because it critiques Reddit, must manually allow comments (WRD)
sies		Politicians and mainstream media
		Passionate group of moderators, topic is their hobby
s and understand the rules and culture		Transparency is also highly dependent on how close the user are with mods
		Regulars are identifiable
		"I'm one of them", open to feedback, user-centric
		Admins don't care and want to make money
		Subreddit inbetwen relations e.g. r/Stellaris and r/Paradox
		Handle insults of users, moderator targeted abuse
		Users type "1984" when their content is removed
		Implicit preference, bar actual real world threats, vile content and extreme levels of hateful content
		Moderators miply users have no responsibility

Figure 4 D.

Selective	Axial	Open
Controversial Content	Derogatory language	Bad actors/User with bad intentions
	Volume	Illegal content, abide by legal framework
	responsibility	Flawed Reddit structure, double standards
it)	Levels	Dealing with meme culture
	experience	Critiques of totalitarian states
	Media literacy	dehumanizing language, but only remove under certain circumstances or have Reddit guide mod
	Spaces	Lacking in discussion and context, heavily influenced by voting system
	Misinterpretations	"who's responsible, everyone is anonymous"
		NSFW tag, content is NSFW and everything needs to be reviewed before posting
		Reddit uses NSFW tag for controversial content
		Subreddit content is not argumentative
		build up a tolerance, mental strain
		As long as its not illegal content should stay up
VRD)		People can disagree with ideas and respect each other, even joke about it
		Because there is no minimizing due to information age, media literacy is still perceived as best route
		Dependet on Reddit structure
		Lack of transparency from Reddit
		No concise clarification of Reddit
		If not dealt with, spills over into real life calls of violence and targeted harrassment
		There are subs for NSFW (but not illegal) content
		Content that gives an unusually large reactions, but also dependent on user opinion, does not have to be offensive
		User-focused approach causes controversial space
		"Clickbait"
t		Controversies are profitable to Reddit
		Where controversial content is, activists are unavoidable, such actions and reactions and situations are unavoidable
		Mainstream media highly impacts what is iseen as controversial or not and makes people dismiss nuance on topics (I
		There exists content that is controversial by nature such as Chris-chan
		Can have mental health repercussions
		Controversial content draws eyes, and in turn clicks which is an incentive from a capitalist structure
		discussed controversial content is almost inherently political
		Often a result of people feeling insulted
		People have different expectations of what should be banned as controversial content, hence why tolerance is importa

Figure 5 D.

Selective	Axial	Open
Human-Bot moderation collaboration	Auto-moderator Experience	"For me its kind of hard to get them working"
	AI moderation	Things done manually instead
	Subreddit size	Used because content can be too much to handle
	Moderator knowledge	Moderation is about power, which is why human interaction is preferred and auto-moderator contrubution is limited
	AI or Algorithm influence	AI have a bad sense of nuance
	AI bot	combination of auutomated tools, bots and large multi-moderators
	Alogorithm manipulation	Moderation should be pimarily a human driven activity
		Default sorting method
		Up to discretion of how moderators set it up, either fruitful discussion sorting, or echo chamber
		Mods lack sufficient knowledge to use efficiently
		Limited use of auto-moderator as the program may cause more problems instead
		Preferred as a support role i.e. Keyword identification
		AI and robots can be biased due to human encoded biases
		Users can influence AI too much
		Bot is really dumb, lacks sophistication
		Tech moderator
		Pushed within larger subreddits instead of human moderators due to shortage
		Used for concrete content such as filtering slurs
		Determines content and in some cases is thus inherently political
		Algorithm is a blackbox
		Clicks = internet loves drame = demand = eyes = money
		Mods do not care ahout automoderator due to limited interaction, trust in moderator community and network for functioning
		Most mods only know the very basics of automoderator
		Inevitability

Figure 6 D.

Selective	Axial	Open
Ethics and philosphical framework	Impossible to standardize	"Impossible meeting ground" (lived experience of moderator 1)
	Free Speech	"Should be a wide degree of allowed content"
	Subreddit culture	"Benefit of doubt should be given to users"
	Legal allegiance	Users flagging content
	Platform/Admin control	mods should be active participants
	Moderation approach	"Reddit must erect a legal team... ...that follows the guidelines"
	tribalism	Exervise editorial decision
	Ideology	violates values of owners
	Mainstream media Skepticism	government hold liable
	Motivation	don't agree with standardization as it would be intrusive and an administrative nightmare
		International disputes, regulatory framework of EU law
		Fails as a symptom of too much information
		Information abundance
		Free speech absolutist, "Its not illegal to be an idiot"
		Slipperly slope if regulatory interferance occurs
		Healthy to be skeptical of mainstream media because contradictions can easily be found
		Political topics
		If it affects people in the real world, that is where the limit is
		Passion and ideology, when these are added to the moderating online space it is inevitable that content will be politically motivated
		Joined for intellectual reasons
		Appraisal of diversity of thought

Figure 7 D.

Selective	Axial	Open
Reddit Ideology	Lack of transparency	A place of contradictions
	Reddit's sustainability	"Reddit is dying but also Buoyant" due to younger generations increasingly engaging with technology but bad moderation practices turn audiences off of the website
	Internet Democracy	Weakness of Reddit versatility
	User autonomy	Users say and feelings on the structure of Reddit
	Fear culture	Change has to occur form the top, Reddit needs to be more clear cut
		Capitalist strcuture, Reddit is a business
		Lacking sufficient options for users to fiter content themselves
		"User third party"
		Companies focused on avoiding bad PR
		Opinions of echo chamber bolstered due to real time voting incentive
		Voting system can be a positive in subreddits as it navigates users to the most educated and more sound answers
		Auto hidden comments if downvoted too harshly
		User can decide how much they want to see or interact
		Lack of recruitment process increases chances of bad actors in positions of moderators affecting healthy discussion
		Mods left to their own devices too much
		Blamed on capitalism, causes the structure on Reddit to survive on a profit and growth incentive
		No one knows what the people in power are doing
		Inelastic spaces, not because its wanted but because its needed
		Controversial subreddits have a more qualified understanding of discussion topic that is controversial
		Reddit is a business first and a social platform second
		Reddit is decentralized
		Profit driven private companies
		Lack of communication, evasive responses and scandals (challenor case)
		No pay for mods

Figure 8 D.

Selective	Axial	Open
Reddit Moderator Culture	Standards	Lack of enforcement by admins
	Power	Time constraints
	Recruitment	Flexible work schedule
		Around work balance and subreddit size
		Mods should be held to a higher standard
		Ideology of moderator matters in moderating, it determines moderating style in the space they are given
		Able to successfully moderate because not a lot is going on in real life
		Reddit mods on Vice documentary, filthy dirty houses, people can't take care of themselves are in charge of large discussion and topical subreddits
		dedication and passion type of promoted job
		Can shut down, and have shut down entire threads
		Applied for position
		Activists incentivized for bigger subs not for passions, which is more for niche subreddits
		Lack of people that want to dedicate their time moderating = anyone = Power mods
		"Most mods are volunteers"
		Part of a team to reduce content over time
		Clear rules to more quickly and justifiably mod content, issue with vague rules
		Easily able to achieve power
		Incentive to "just get the job done" instead of meaningful moderation
		Mods interpret content among themselves with only their interpretation to check the content and provide to admins
		Bad actors make more of a difference
		If people don't like moderation culture, they can apply themselves
		reddit's reputation of moderation culture is conflated with all moderators
		Activists with political motives
		Its an ecosystem
		Lack distinction from real world, lack connection to reality, face to face contact
		"We have to express our desires, spread that message"
		"Job can be stringent"

Figure 9 D.

Selective	Axial	Open
Incinserity	hypocrisy	Flawed Reddit structure and double standards
	Corruption	Mods using powers for their own ends
	Double Standards	Tribalism is more identifiable on the internet
		Banning content perceived as unfair
		Reddit Admin Watch, double standards present on watch dog subs
		Nepotism, Passionate, Ideological
		Power is more of a mindset, disconnected and disillusioned
dits		Absolute power corrupts absolutely
		Sm,all minority that makes decisions for everyone
		People don't understand how these tools can be used against them
		Concerned with banning and censoring info that can harm their image
		Their inaction and allowance of bans causes spaces where everuone agrees and ideas aren't challenged
		Using of CP to shut down other boards
		AHS allowed up, even if they violate rules which break the law
		Blatant double standard
		Reddit lenient with stuff they agree with, Us vs. Them mentality when conducting operations
		Passionate leads to biases?
		subreddit double standards
		Social immunity "us vs. them" "Nothing I can do"
		"Authoritarian" reputation
		Fear, peopl are afraid how other subreddits may react
		Sense of entitlement and think they're above userbase
		Think of it as a priviledge, not a responsbility
		r/Popular shows the large scale misinterpreting actors that lack nuance (2balkan4u)
		Do one thing, say another
		Other subreddits address issue rather than Reddit because its easier
		Insincere moderation that is explicit, that is even well-known and even accepted standard
		"Some mods just don't care about transparency"
		Dellusions of Granduer, Some mods see themselves as saviours of democracy, justify their moderation practices and attitutdes accorting to their b
		self-centered attitudes
		Certain moderators on Reddit have questionable moderation style

Figure 10 D.

Selective	Axial	Open
Social commentary	Society Problem runs	Problems of society on mental health run deep and users use social media platforms as their outlet
	situational	Rather than solving real world problems it's a case of slacktivism
		Focus on virtue signalling rather than solving real issues
		PR focus due to capitalist incentive causes for no actual progress and the devation of real world problems
		Commentary on history, politics, right wing and left wing politics, identity politics
		Public outcry is the only way to make coporation care
		social media attention
		Platform of users starting to complain is the only way to bring about change
		Board members and executives that are out of touch only care if outcry is had
		Global platform makes it impossible to moderate effectively
		Controversy is unavoidable

Appendix E.

Consent form

Figure 1 E.

INFORMED CONSENT FORM

Project Title and version	How do Reddit moderators perceive the structure, motivators and actors of the Reddit ecosystem that influence their moderation style and community surroundings?
Name of Principal Investigator	Maarten van der Stad
Name of Organisation	Erasmus University
Name of Sponsor	
Purpose of the Study	This research is being conducted on the decisions made by moderators on controversial content, controversial referring to not entirely illegal or against the rules, but may be provocative or challenging. I am inviting you to participate in this research project about this topic. The purpose of this research project is to collect data and gain a nuanced understanding of the moderator perception of the topic.
Procedures	You will participate in an interview lasting approximately 40 minutes. You will be asked questions about moderation, content you kept and removed deemed controversial by the subreddit, the suggested moderation rules by reddit, and harrassment you may have experienced on your end as a result of moderation of content, and the perception of the user on the moderator. The interview will be recorded and your input will be used anonymously in my research. The recording will be deleted after grading. You must be a moderator on Reddit, and be an active moderator in at least 1 subreddit for the past 3 months or more.
Potential and anticipated Risks and Discomforts	There are no obvious physical, legal or economic risks associated with participating in this study. You do not have to answer any questions you do not wish to answer. Your participation is voluntary and you are free to discontinue your participation at any time.

Potential Benefits	Participation in this study does not guarantee any beneficial results to you. As a result of participating you may better understand how you perceive moderation processes, to what standards controversial content can or should be defined (possibly by law) and what changes need to happen for online moderators.
Sharing the results	Our research paper will be available from October 20, 2022 onwards. Upon request, it will be sent to your email.
Confidentiality	Your privacy will be protected to the maximum extent allowable by law. No personally identifiable information will be reported in any research product. Moreover, only trained research staff will have access to your responses. Within these restrictions, results of this study will be made available to you upon request. As indicated above, this research project involves making audio recordings of interviews with you. Transcribed segments from the audio recordings may be used in published forms (e.g., journal articles and book chapters). In the case of publication, pseudonyms will be used. The audio recordings, forms, and other documents created or collected as part of this study will be stored in a secure location in the researchers' offices or on the researchers password-protected computers and will be destroyed within ten years of the initiation of the study.
Compensation	*N/A*
Right to Withdraw and Questions	Your participation in this research is completely voluntary. You may choose not to take part at all. If you decide to participate in this research, you may stop participating at any time. If you decide not to participate in this study or if you stop participating at any time, you will not be penalised or lose any benefits to which you otherwise qualify. If you decide to stop taking part in the study, if you have questions, concerns, or complaints, or if you need to report an injury related to the research, please contact the primary investigator: 474033mv@student.eur.nl

Statement of Consent	Your signature indicates that you are at least 18 years of age; you have read this consent form or have had it read to you; your questions have been answered to your satisfaction and you voluntarily agree that you will participate in this research study. You will receive a copy of this signed consent form. I have been given the guarantee that this research project has been reviewed and approved by the ESHCC Ethics Review Committee. For research problems or any other question regarding the research project, the Data Protection Officer of Erasmus University, Marlon Domingus, MA (fg@eur.nl) If you agree to participate, please sign your name below.	
Audio recording (if applicable)	I consent to have my interview audio recorded ☐ yes ☐ no	
Secondary use (if applicable)	I consent to have the anonymised data be used for secondary analysis ☐ yes ☐ no	
Signature and Date	**NAME PARTICIPANT**	**NAME PRINCIPAL INVESTIGATOR** Maarten van der Stad
	SIGNATURE	**SIGNATURE** 
	DATE	**DATE**

Appendix F.

Figure F 1. Interview Guide/Topic List

Topic List

Probing questions

- **Introductory questions**
    - How long have you been a moderator for?
    - Why did you decide to take up the job of a moderator?
    - Does it become tiring at points, or is it actually invigorating to work with this type of content, or better yet to moderate it?
    - How fundamentally flawed is Reddit and its structure really, and why do you believe it is shaped this way?
- **Your role as a moderator**
    - What do you believe your role is for on the internet forum on Reddit?
    - Are you aware of the different processes of moderation, and if so which process do you identify yourself with, and which process do you think most other moderators identify with?
    - How do you think you are perceived by the average user in your subreddit?
- **Controversial Content**
    - How would you define controversial content, and when confronted with it, what is your first response?
    - How familiar are you with official laws on hate speech and what is your interpretation of applying them?
- **Moderator tools**
    - Are tools more implicit oriented or explicit oriented?
    - Do you use the automoderator and bots to filter content, if so or if not, why?
    - What parts of the tools do you most use and for what purposes?
- **Time contraints**
    - How do you deal with the ratio of content to your ability to moderate it?
    - Does it depend on style of moderation, or on other existing factors?
- **Micellaneous**
    - Would you like to add something else?
    - How did you feel about X case study

Appendix G.

Fragments of Coded Interview

Figure G 1.

[00:05:32] **Reddit Moderator Interviewee 9:** Well, I'd say that I used my moderation style. Yeah. People know what we're doing. (Interviewer mumbles "yeah" during speech) Um, but we don't really try and take over everything and moderate every discussion. A lot of the times where it feels like we're a little bit, you know, invisible hand, like beating that crowd,, just changing a few things every now and again. But I feel that, I mean, [00:06:00] yeah, like our rules and the way that we moderate, right. Isn't too oppressive or like dictatorial, um, Yeah, we have a very casual relationship, with our userbase and that..

Figure G 2.

[00:13:32] **Reddit moderator interviewee 3:** Yeah. I mean, I agree that maybe these companies can do bad and communication of letting them, like letting people know where their beliefs lie. A lot of the times you have a lack of communication. Why did these people get banned? Why did this happen? And then it's just radio silence, or very, you know, uh, uh, like evasive response when you know, where they don't really answer to the question. I think companies could do a lot better job of actually. Uh, [00:14:00] clarifying where like, what are they going to do? How they going to do it to all of a sudden content? I mean, yeah, like I said, the bias is implicit. It can't be removed because obviously these are private companies. These are not government entities that have a sole intention and even governments (interviewer mutters "yeah" during speech) have certain biases, especially in, you know, non EU countries or countries where information is restricted. (interviewer mutters "yeah" during speech) Um, yeah, it's just. It's kind of hard, especially with, I'll say private companies and telling them them what to do. (interviewer mutters "yeah" during speech) Cause obviously we know what power they have and we know a lot, it takes a lot of effort and a lot of money, especially that's what matters a lot of money into getting to the, like getting them to do what you want.

Figure G 3.

[00:05:25] **Reddit Moderator Interview 4:** You know, so I, we are in restrictions, in which other subreddits aren't, (interviewer mutters "yeah"during speech) which makes it all sort of like, from user perspective. I'm sure that there's a bunch of like hypocrisy of in this sense that like, you know, you have to manually allow most comments and stuff like that.

[00:05:51] **Interviewer:** Yeah.

[00:05:53] **Reddit Moderator Interview 4:** So it's like very sort of counter productive in terms of allowing discussion.