

The Influence on Attitude towards Piracy: a Sentiment Analysis of Dutch Media Outlets

Name: Marieke Paardekoper

Student Number: 0848913

Supervisors: Max van Duijn & Jaap van den Herik

Course: Graduation Project Media Technology

University: Universiteit Leiden

Date: 29-07-2018

Abstract

Intellectual property piracy is illegal and costs the relevant industries significant amounts of money, yet many people download without remorse. This attitude towards piracy has been one of the reasons why it has been difficult to battle. Through the use of the Social Cognitive Theory it can be found that two aspects influence the attitude of pirates: outcome expectancies and social learning. Outcome expectancies influence people towards piracy, since there are many rewards and little risks or sanctions. Social learning's effect is significant, but it is harder to determine whether it is positive or negative, since social learning concerns the influence of peers, and in particular that of media. By performing a sentiment analysis on Dutch news articles, the attitude of media is tested to find out whether social learning is a positive or negative influence. From the results it can be concluded that media do not deter people from downloading and are mostly neutral in their portrayal of piracy. This means that a possibility for changing attitude towards piracy might be found through the use of media.

Contents

1. Introduction	5
1.1 My Motivation.....	5
1.2 Problem Statement and Research Questions.....	6
1.3 Structure of the Thesis	6
2. Theoretical Background	7
2.1 A Short History of Piracy.....	8
2.1.1 The Growth of Peer-to-Peer	9
2.1.2 The BitTorrent Protocol	10
2.1.3 Other Methods of Downloading.....	10
2.1.4 Prevention of Piracy.....	11
2.2 SCT: Outcome Expectancies	12
2.2.1 Copyright Laws in the Netherlands	12
2.2.2 Implementation of the Download Ban	14
2.2.3 The Effects of Litigation.....	16
2.3 SCT: Social Learning	16
2.3.1 Commercial Prevention	17
2.3.2 Media Influence and an Answer to Research Question 1	19
3. Sentiment Analysis: Method	20
3.1 Defining and Explaining Key Components of Research Question 2.....	20
3.1.1 What is Media Depiction?	20
3.1.2 What is Intellectual Property Piracy?	21
3.1.3 Why This Timeframe?	22
3.2 The Corpus.....	22
3.2.1 Creating the Corpus.....	22
3.2.2 The Algorithm.....	25
4. Sentiment Analysis: Results	29
4.1 The 'Illegal' Issue.....	29

4.2 Results29

4.3 Overview of the Data: an Answer to RQ234

5. Conclusion & Discussion36

5.1. Summary of the Results36

5.2. Context of the Results and a Solution to the Problem Statement37

5.3 Further Research38

6. References39

1. Introduction

I have three confessions to make:

- I used to be a pirate.
- Many of my friends used to be pirates and there might be a serious possibility that they still are.
- I feel no particular shame for my past, nor would I seriously condemn my friends if they are still committing piracy.

I am, of course, talking about digital piracy which might make those confessions a bit less drastic. But, even though digital piracy is not quite on the same level of crime as actual piracy, it is still a crime. In fact, piracy is an enormous problem for anyone wanting to protect their intellectual property as someone else could download their property without paying for it. Streaming services such as Spotify and Netflix were expected to drastically change the amount of data that is illegally downloaded annually. However, these results have been disappointing – as can be seen from the numbers below:

“Approximately 99% of data transferred on peer-to-peer networks is copyrighted, 42% of the software currently in use worldwide is pirated, more than 75% of computers have at least one illegally downloaded application, 95% of music downloaded online is illegal (the rate in the United States alone is 63%), 66% of online torrents are illegal, and 22% of Internet bandwidth worldwide is used for piracy (Go-Gulf¹).”

1.1 My Motivation

In addition, it is not just the amount of data that gets downloaded. There is a significant number of people who download, as is shown in an investigation performed by CentERdata and IVIR, where 25% of respondents indicated that they download illegal material (Poort & Leenheer, 2012). How is it possible that so many people download, and why do I simply accept my (and that of my friends’) history as a pirate without any remorse? One could argue that this is a cultural problem, that I grew up surrounded by friends who all committed the same crime and therefore normalized it, but that still would not explain the high piracy rate. And even now, when discussing piracy with my colleagues who surely grew up surrounded by different friends, piracy is talked about as if it were normal. In fact, it turns out that 70% of all internet users find that there is nothing wrong with piracy (Lowry, 2017). This acceptance of criminal behavior baffles and intrigues me, and when

¹ <https://www.go-gulf.com/blog/online-piracy/>

trying to find a reason for it, there was only one common denominator that I could think of: media influence. If most people share the same view on something than it must come from somewhere, which is why I decided to research this influence.

1.2 Problem Statement and Research Questions

To execute the research, I have posed a problem statement that will serve as the backbone of the paper. To find the reasons where the problem came from and why it persists, I will first form a theoretical background to find out what external factors influence people to commit piracy with a special focus on the influence of media. After this, I will determine the light in which media depict piracy by performing a sentiment analysis on 60 different articles within a confined timeline. In an overview, it looks like this:

Problem statement: Intellectual property piracy is illegal and costs the relevant industries significant amounts of money, yet many people download without remorse.

Research question 1: Which external factors influence people to commit piracy, and do they do this in a positive or negative way?

Research question 2: Was the media depiction of intellectual property piracy in Dutch written media positive or negative between 2009 and 2014?

1.3 Structure of the Thesis

The research questions will be discussed per chapter to provide clear and consecutive answers. In Chapter 2, research question 1 will be answered by providing a framework that determines the reasoning for people to start downloading. This framework consists of five categories, of which two will be further discussed to determine their influence on the piracy behavior. From this we may conclude that media influences people in their behavior towards piracy. Following Chapter 2, Chapter 3 and 4 will handle research question 2 by performing a sentiment analysis on 60 news articles written between 2009 and 2014, taken from 6 Dutch written media outlets. Chapter 3 will clarify the method that was used to perform the analysis, Chapter 4 will detail the results of the analysis. Finally, Chapter 5 will summarize the answers to research questions 1 and 2 and use them to find a reason why the problem in the problem statement still persists and to what extent a possible solution exists. This chapter will also consist of a section to induce further research.

By performing this research, I hope to shed light on piracy itself, the culture surrounding it, and the ways to fight it in the future.

2. Theoretical Background

Since digital piracy is a large problem, it comes as no surprise that extensive research has been performed into the issue. Since 70% of pirates finds that there is nothing wrong with piracy (Go-Gulf, 2011; RIAA, 2015), many researchers have focused on the disconnection between this attitude and the negative consequences of piracy. Many theories have been applied in various manners, such as deterrence theory and self-control theory, resulting in contradictory findings and differing emphases and conclusions. Finding a consensus within this field has therefore been difficult, resulting in uncertainty towards the actions to be taken against piracy. A recent meta-analysis of 257 empirical studies focused on digital piracy (Lowry et al., 2017) aimed to take away this uncertainty by finding “unifying answers to advance the research and practice associated with preventing the noxious global problem of piracy”. Within this meta-analysis, the Social Cognitive Theory (Bandura, 1986) came forward as being the most cited option for a theory that can set up a framework explaining piracy behavior.

Social Cognitive Theory (SCT) builds on the idea that criminal behavior is learned by watching others, but it also adds that criminal behavior is influenced by social and environmental factors, such as psychological outcome expectancy determinants, environmental determinants, observational learning, and self-regulation. SCT further emphasizes reciprocal determinism, which is the idea that personal factors (e.g., self-efficacy), behavioral factors (e.g., positive/negative responses to behaviors), and environmental factors (e.g., facilitating conditions) affect each other reciprocally. Behaviors and their associated consequences interact further with personal and environmental factors in the reinforcement process, in which people learn to repeat beneficial behaviors and to avoid harmful ones (Bandura, 1986; Higgins, 2007; Glanz, Rimer, & Viswanath, 2008). Lowry categorizes SCT in the following five major categories:

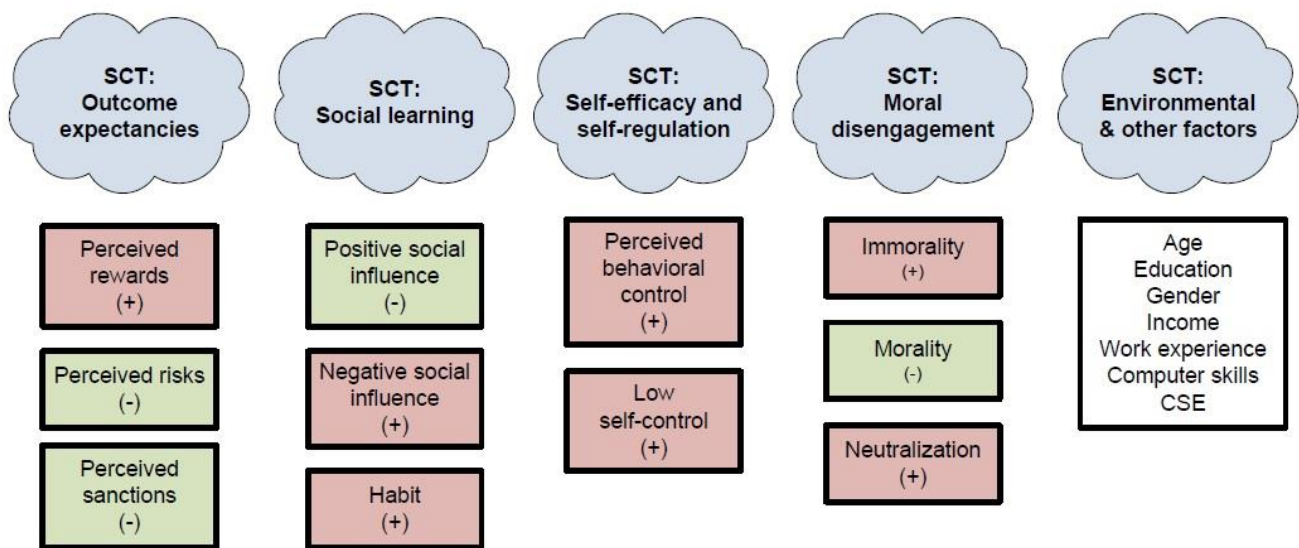


Figure 1: SCT-based framework of the major predictors of digital piracy (Lowry et al., 2017)

The rectangles in red signify a negative influence on preventing piracy, while the green convey a positive influence. For instance, perceived sanctions are an outcome expectancy and they prevent people from committing digital piracy, while perceived rewards do the opposite and drive people to illegally download. For Environmental & other factors, elements like CSE (computer self-efficacy) or age can either positively or negatively influence someone depending on the specifics for individuals. All these categories work together to affect someone towards either a negative or positive attitude with regards to piracy.

The purpose of this chapter is to answer the following question: which external factors influence people to commit piracy, and do they do this in a positive or negative way? To answer this question and to understand how external factors have influenced people, we will take a look at the specifics for the first two SCT categories: outcome expectancies and social learning. Self-efficacy and moral disengagement are not included in this overview, as they are almost completely focused on individuals and their own (internal) motivation. Since we aim to determine the influence of the external factors in illegal downloading, these internal motivations are not of interest. Environmental factors are specific for individuals and the stimulants that someone might have within a social group. While this is influenced by external factors, these are not as easily determined as either being positive or negative towards piracy. Additionally, if several environmental factors are of import for an individual then it becomes unclear which factor will have the most influence. Without an extensive research focused on individual choices and influences, it is therefore impossible to precisely determine the impact of the factors. For this reason, these factors are out of scope for the research and will not be included within the framework.

To understand how outcome expectancies and social learning affect pirates, we will review the influence the factors have had in the past and how this has resulted in high piracy rates. To provide a framework for these, I have first detailed a short history of digital piracy itself in Section 2.1. The outcome expectancies factor is handled in Section 2.2, where the history of Dutch litigation against piracy is detailed and reviewed on how this has influenced pirates. In Section 2.3, the social influence is factor discussed, which focuses on the influence of peers as well as media.

2.1 A Short History of Piracy

In 1999, a 19-year old student named Shawn Fanning developed a piece of software that would enable users to easily copy music files from each other's computers (the so-called *peer-to-peer* model, or, *the client-server model*). This was the start of Napster and it would not take long for the program to become a thorn in the side of copyright owners around the world (Honigsberg, 2002). The files distributed through Napster were stored on the local hard disks of the users, and not on Napster's servers, which only provided an index to all files that were available. To download a file, a connection was established between the downloader and the uploader (i.e., the provider of the file). Downloading the file then happened through this connection - Napster was not

directly involved, but did facilitate the connection between the users. To do this, the program kept a list of every user who was currently online and which connections they had made. Such a list in the current environment would make it much easier to find pirates, but during the early 2000's, piracy had not quite been evaluated in the way that it is now. Pirates were therefore not sought out, and Napster's popularity continued to grow. Between 1999 and 2001, 20 million people made use of Napster and the program was downloaded approximately 80 million times (Freeman, 2008). Record label companies were now quick to realize that this program would cost them money, and consequently took action against it. The largest record companies in the United States came together to sue Napster for copyright infringement, and in 2001 the court decided that Napster could be held liable for copyright infringement (*Metallica vs. Napster*, 2001). Napster was fined generously, and the company filed for bankruptcy in 2002. In addition, digital piracy was now officially declared illegal.

2.1.1 The Growth of Peer-to-Peer

Despite the shutdown of Napster, peer-to-peer (P2P) piracy continued to gain users. The music industry had won the battle by winning in court from Napster, but they had not won the war – in very quick succession after Napster was taken down, software like KaZaa, BearShare, and LimeWire popped up to provide the users with different ways of P2P sharing of files (Choi, 2006). However, these programs, unlike Napster, allow for the exchange of files other than music. People now had access to programs that would easily allow them to pirate music, movies, and software. In addition, the creators of these programs abandoned certain methods that made it possible for Napster to be held accountable for the copyright infringement. The servers no longer contained a list of which user provided which file and the search request became decentralized. Instead, certain software started using the computers as junctions (so-called supernodes), which provided a temporary index of all the files that were currently being offered by users. The request could also be passed along to other supernodes, ensuring that every file being offered could actually be found – and that there was not a central location from which every user (and their files) could be located. Users now enjoyed more anonymity, as it became increasingly difficult for organizations to find the pirates; tracking data usage may have given an indication as to someone's downloading or uploading habits, but this data was not public and only available to the Internet Service Provider (ISP) who would not be interested in sharing this information to protect the privacy of its users.

2.1.2 The BitTorrent Protocol

As a culmination of all these developments, it was in 2001 that the real successor of the P2P protocol was created: the BitTorrent protocol, which works as follows. Through the use of a BitTorrent client (like µtorrent or Azure) someone can open a .torrent file which is hosted on websites like The Pirate Bay and Kickass Torrents. This file contains information on what exactly is to be downloaded, such as the exact size of the file; the name; and whether or not the file is divided into smaller parts. The BitTorrent client will then make a connection with a decentralized server that connects the client with others BitTorrent users (peers) that already have the complete file or are currently downloading it (seeders). The different parts of the file are downloaded from as many peers as possible, and everyone that has downloaded that part to their computer becomes a peer themselves which ensures that everyone will share the file and that it will remain available for other users. Once a user obtains all of the fragments that comprise a file, the file is reassembled in its entirety on the recipient machine. This makes the protocol particularly useful for large files because it most optimally utilizes bandwidth by spreading the file distribution to all individuals who each have varying fragments. Furthermore, the problems of "leeching" - where individuals simply download (leech) from others without "sharing the wealth" is reduced because individuals only gain download speed when they give upload speed to others. This encourages pirates to always share the files that they have downloaded (Hinduja, 2006). Notable about this new way of sharing files is that there is hardly any communication with a central server necessary. The only time a user will have to communicate with this server is to gain access to the .torrent file. Since BitTorrenting does not use a single piece of software to track the users of the BitTorrent protocol it becomes much harder to find any pirates. A consequence of this is that the index sites that provide links to .torrent files are often held reliable for the copyright infringement that is committed through the use of the website (Steinmetz, 2010). While the anonymity of the users offered protection for the pirates, it was the notoriety of the index sites that ultimately proved to be an Achilles heel for the BitTorrent protocol. With the shutdown of several websites like The Pirate Bay and Kickass Torrents, pirates have needed to find different ways of sharing files (Anderson, 2007).

2.1.3 Other Methods of Downloading

Before we explore the newest ways in which pirates shares their files, there is one more popular pirating method that needs to be discussed: the direct-download model (Choi, 2006). This model offers a simple and fast way for file distributors to save files on servers that are hosted by the companies that participate in this model (such as MegaUpload). The user that intends to download a file is given a URL to find the file, and can then immediately download it to their own computer. This means that the downloader is no longer connected with different users. In addition, the hosting companies have learned from their predecessors, and do not provide a search function. Downloading can then only be done when someone is given the URL – making it

far less accessible and therefore less exposed. But when the popularity of the model rose, it was only a matter of time before search bots popped up, making it possible for users to find materials that they wanted to download. With more accessibility, the model became even more popular and it was therefore no surprise that steps were taken against it. In January 2012, MegaUpload was taken down by the United States Department of Justice after the owners were indicted for allegedly creating an organization dedicated to copyright infringement (Yung, 2012). With the loss of the biggest site that utilized the direct-download model, the popularity shrunk and many people found new ways of gaining access to data. For TV-shows and films, one method in particular rose in popularity: streaming.

While streaming sounds different than downloading through the BitTorrent protocol, it is in fact very similar. When looking at the current most popular free streaming service, Popcorn Time, we find that it is simply software that downloads a torrent from an index site. The difference lies in two small changes: people download through use of the software and they can start watching a movie or TV show before the download is complete. Both of these functions do not change the way in which the file is downloaded, but they do make downloading much more accessible for its user. And since Popcorn Time deletes the file once it's been watched, and they have built in a Virtual Private Network (VPN) to ensure that its users can't be tracked, it's also become even more anonymous for people to download.

2.1.4 Prevention of Piracy

It is clear that the technology surrounding piracy has been changing and evolving continuously, and almost always in favor of the digital pirates. Prevention has always been a step behind, at first because piracy was such a new commodity and later because its proven to be very hard to prevent people from downloading when you do not even know who they are. The current prevention methods therefore mostly focus on Digital Rights Management (DRM) which tries to deny access to the software or files instead of preventing people from downloading them. During the era of the DVD this was done by encrypting DVD's in such a way that they could not be copied to an empty disc, but with digital streaming on the rise it is mostly done through watermarking for movies and forcing users to be online for software and games. But watermarking is ineffective in preventing people from watching movies, encryption is ultimately almost always cracked, and forcing people to be online for the use of software prevents paying users without a secure internet connection from using the software and it requires companies to keep servers up – if the servers are ever taken down then the software can no longer be used. The last consequence in particular has been the reason for much uproar about the use of DRM. It therefore remains to be seen how successful the use of DRM will be.

All in all, the technology surrounding piracy has rapidly evolved over the years, and an arms race between pirates and developers of anti-piracy-related software has only increased the speed in which new

technologies are used. With this in mind, let us take a look at some external factors that might influence someone to commit piracy.

2.2 SCT: Outcome Expectancies

Outcome expectancies are the category of SCT that are most influenced by direct actions of governmental institutions. While the perceived rewards will always be the largest reason for pirates to download, both perceived risks and perceived sanctions are consequences of legislation and law enforcement. If these are both restrictive, then one could expect that more people are deterred from piracy. But one of the biggest issues with the rise of piracy in the Netherlands has been the uncertainty of what was allowed within the confines of the law and what was illegal. The rise of new technology has clearly been swift, creating quite some difficulty for authoritative sources to make laws that can keep up with all changes that occur, which in its own right generated widespread confusion and legal uncertainty (Depoorter, 2006). By taking a look at the history of legislation and their effects, we can estimate the impact of the outcome expectancies in the influence of committing piracy. This will be done by detailing copyright laws in the Netherlands in Section 2.2.1, after which the focus will shift specifically to the implementation of the download ban in Section 2.2.2., which was the Dutch law that made downloading copyrighted material illegal. In Section 2.2.3 the effect of the methods of litigation are highlighted to show the influence it might have had on pirates.

2.2.1 Copyright Laws in the Netherlands

Software piracy is a well-known term that indicates the unauthorized use or reproduction of another's work. In practice, the meaning of this is less easily determined since it is hard to clearly define what exactly makes a work the property of someone. In the Netherlands, the rules that signify such things are written down in the 'Auteurswet'. This law can broadly be described as follows:

“The auteurswet is the exclusive right of the maker of a work of literature, science, or art, or their legal successors, to make it public and to reproduce it, barring the restrictions, as stipulated by the law.” (art. 1 Wet AW 1912)

The first right of the work is assigned to the creator of the work, and they obtain it the minute that the creation of the work commenced. This last part is important, because it indicates that the right arises without any interference, meaning that the law of the author is obtained without the need for formalities. For example, anyone who has created a painting at home owns the rights of the work, simply on the basis that they have started it. Of course, the work has to be sufficiently original for it to apply.

The right of the author will not stay with the original creator of the work indefinitely; it can be taken over by someone else - for instance when the creator sells the right. When someone has the ownership of a work, he or she is owner of two exclusive rights: they can publish the work and they can reproduce it. A consequence of the existence of these rights, is that everyone who wants to publish or reproduce a work must have permission from the current owner of these rights. If an owner of a copyright dies without signing the rights away, the work remains protected for another 70 years, after which it enters the public domain and can freely be used by anyone who wishes to do so. Publishing a work by offering it on an index site like The Pirate Bay is therefore clearly not in accordance with what is allowed in the Netherlands. Downloading from such a site is not the same as republishing a work, and it therefore doesn't violate this law, though it does violate the so-called homecopy law. This law has been the subject of many changes in the last decade, and has been the cause of quite some controversy.

The homecopy law, which was created in 2003, concerns itself with everything that surrounds the copying of copyrighted material – and it also specifically states the circumstances under which someone is allowed to do so. These circumstances are as follows:

1. The copy is not a, direct or indirect, means for monetary gain;
2. The copy serves exclusive to own practice, study, or use;
3. The number of copies are limited, or the creator of additional copies compensates the holder.

What is most notable from this list is that anyone can make a copy of a protected work, as long as they do not share it with anyone else. In its basis, this means that downloading can freely be done as long as people do not upload. In favor of this law, it was often argued that uploading is such an integral part of downloading with torrents that this restriction was good enough to protect someone's copyrighted work from being obtained illegally by too many people. The argument that downloading someone's material should not be allowed at all was also considered, but it was often thought that there was no way to prevent it without restricting access to the internet. It was therefore a different aspect of the law that would first be changed: the homecopy tax. This tax is a fee that consumers pay when they purchase all kinds of electronic products that are used to copy material (e.g., empty disks and hard drives). The idea behind this tax is that it can function as a reimbursement for companies that feel like their copyrighted material gets stolen on a regular basis. There were two heavily discussed aspects of this law – the first is that everyone pays this tax, not just the downloaders. The second is that an environment was created in which the owner of a copyright would have to tolerate these home copies since they were already being compensated for it.

Since the companies felt that they would have to be properly compensated to agree with this, there was quite some discussion about the actual height of the tax. A group of electronic companies, including Sony and Philips, objected to this in 2012 (*ACI Adam BV and Others vs. Stichting de ThuisKopie*, 2012), as they felt that it negatively impacted their sales and that it was unfair that non-downloaders paid the cost for what

others did. The companies ultimately went to court over this, and it reached the European Court of Justice in 2014. They took a new look at the homecopy law and ruled it to be an infringement on a European directive (2001/29/EG – article 5 § 2-b and § 5). The court indicated that it was unacceptable that the homecopy law does not make a distinction between copying from a CD someone bought and downloading copyrighted material. The first result of this was that the homecopy tax was lowered. A second result of this verdict was that another Dutch law was passed, called the ‘download verbod’ (download ban), to serve as addition to the homecopy law. This law, which was passed in 2014, finally made it illegal for people to download copyrighted material from an illegal source even when they do not share it with anyone else.

2.2.2 Implementation of the Download Ban

Though the ruling from the European Court did not come as a real surprise, the decision still created an uproar of discussion about one subject in particular: privacy. With the download ban in effect, it would mean that people could (at least theoretically) be prosecuted for downloading – but how would the execution of this be done? The answer to this has to be found with the Internet Service Providers (ISPs), which are the companies that offer users access to the internet. These ISPs are an interesting phenomenon when looking at piracy from a legal and economic perspective (Elkin-Koren, 2006). These are, after all, companies that could be held responsible for the losses that property piracy causes – as they benefit from the illegal activities that are made possible by the services that they offer and they have the power to oversee them. They are also easier to identify than individual users would be. Though, of course, expecting the ISPs to solely be held responsible for what its users do is dangerous and undesirable; it would require them to police its users and make restrictions on basic rights like having access to the internet. This is why the law has shielded them from this by offering them restrictions on this liability if they exercise control over the online distribution of illegal copies of copyrighted materials. A copyright owner could therefore ask an ISP to stop showing a website that offers pirated material and it would have to adhere to this request.

The only other actors who can then be held responsible are the downloaders themselves, but to get to them, the ISPs still play a central role. They have the information that is necessary to determine the location of the user (their IP address), but they will usually only give this information if a judge asks for it – they are obligated to do this, privacy laws do not protect the users in this case – or if they find themselves that a user is conducting criminal behavior. The latter case requires the companies to search for a crime without a real incentive for them to do so and it is therefore no surprise that this hardly happens. In addition, there are ways for users to hide their IP address (most notably through the use of a VPN), which makes it almost impossible to catch them unless more technological advances are made. Perhaps this is the reason, that it is currently not even the intention of the download ban to litigate.

Even though the law can penalize people through criminal law, the public prosecutor drafted policy papers that stated how they will handle the transgressions of this law. It can be concluded from the papers that it was not the intention to have people be incarcerated for downloading copyrighted material and in practice this has turned out to be true. However, someone can be held accountable through the Civic Court for downloading. Someone who holds a copyright can indicate a grievance with the judge that their copyrighted material was used in ways that are not in accordance with the law. In the Netherlands, this means that Stichting Brein, which is the non-profit foundation that manages the copyright complaints, will be able to press charges. Currently, they are mostly looking for people and websites that provide large amounts of copyrighted material through torrents. In January 2016 they announced that they intent to charge more people who do this, for which they developed new software that should be able to help them with this cause. They have never charged someone for simply downloading; the responsibility for that currently lies with the copyright holder.

But it is not just the Dutch court that can sue a downloader from the Netherlands, as jurisdiction on the internet is a tricky and fickle mistress. On principle, the laws from the country that the user downloads from are the ones that apply – the country that houses the servers does not necessarily play a role in this. However, this does not always turn out to be true. A perfect example of this is the previously mentioned case against MegaUpload. In this case, the USA turned out to be willing to push the boundaries of the jurisdictional laws, simply because of the interests at stake. The offices of MegaUpload were stationed in Hong Kong and New Zealand, but a large amount of the servers were located in the USA. The financial transactions of the site were also largely conducted through the use of American companies (like PayPal and Google AdWords), which – according to the charge – was the reason that MegaUpload knew that it was doing business in America and therefore exposed itself to the American judicial system. It therefore was on the USA federal prosecutor's request that the offices in New Zealand were raided in 2012 and the request was made to have the creator of MegaUpload be extradited to the USA. It has been the USA in general that has taken the greatest steps against piracy through jurisdiction, with the use of statutory damages being the one that jumps out. This stipulates that the infringer of the copyright is held liable for (1) the copyright owner's actual damages and any additional profits of the infringer, or (2) statutory damages (U.S. Code § 504). There are two aspects of this that make it a controversial law. The first of this is the fact that the USA lets the owner of the copyright that was violated decide the cost of the infringement, whether those were factual damages or intended ones. This means that the American system lets the plaintiff choose to lower the burden of proof, as they only need to show that the defendant infringed the copyright, and not how that financially impacted the plaintiff or if it even had an impact at all. The second is that the law stipulates that a granted statutory restitution for every infringement with regards to every single work, can't be less than 750 dollars or more than 30.000 dollars. This creates the possibility for enormous discrepancies between different cases of copyright infringement even when they appear to be very similar (Samuelson, 2009). It is also this ability to inconsistently have excessive fines for piracy-related crimes that creates such a hateful relationship between the industries that are established in the USA and some of its consumers.

2.2.3 The Effects of Litigation

This section has sought to answer how outcome expectancies influence people to commit piracy and whether they do this in a positive or a negative way. From the history of piracy and litigation in the Netherlands, it has become clear that perceived rewards easily outweigh the perceived risks and perceived sanctions. Anyone deciding on whether or not to commit digital piracy will hardly be deterred by any sanctions as none that are currently implemented are of any real consequence. Though laws have become much stricter in the last few years, the effect on piracy has been minimal. It is clear that the litigation strategy that is currently maintained will probably not deliver the desired result of a decrease in piracy. In addition, with as many piracy index sites that exist, it becomes very easy for anyone to find the file that they are looking for. This makes the rewards easily achievable. Outcome expectancies therefore exert negative influence by moving people towards illegal downloading more than they deter them from it.

2.3 SCT: Social Learning

There are three factors mentioned in the SCT that can influence piracy behavior within the social learning category: positive social influence, negative social influence, and habit. It is the last one of these three that seems the hardest to determine the influence of. People got in the habit of downloading when there was judicial uncertainty and everything was new and exciting; only improvement of other SCT categories can possibly get them out of this habit. The influence of habit is therefore large, but it is difficult to determine in what way it outweighs the other SCT factors when it comes to determining to commit piracy. When looking at negative and positive social influence, we find much more clear factors.

One of the most well-known theories concerning the social influences on unethical behavior is the social learning theory (Bandura, 1977). This theory, which SCT is derived from, proposed several ways in which a behavior is learned:

1. Learning is not purely behavioral; rather, it is a cognitive process that takes place in a social context.
2. Learning can occur by observing a behavior and by observing the consequences of the behavior.
3. Learning involves observation, extraction of information from those observations, and making decisions about the performance of the behavior.

The social learning theory can then be applied to find that people started downloading through the influence of peers. This was also concluded by Hassan Aleassa (2009), who investigated the downloading behavior of students and found that community acceptance influenced the choices of the students:

“We can conclude that [the] students’ intentions to pirate software are shaped mainly by social influences and secondarily by their attitudes toward software piracy. [...] We found that subjective norms had a stronger effect on intention to pirate software relative to attitude. This means that students’ intentions were greatly influenced by opinions and behaviors of significant others.”

These ways of learning new behavior can originate from any association with others, including the media. Of the influences that a person encounters on a daily basis, it is the media that is most easily measurable in its outspokenness towards piracy. As the media is a public domain, it has already been used to try and dissuade pirates from committing intellectual property theft. This has mostly been done through the use of campaigns and by emphasizing publicly that piracy is stealing. The method has had little success, however, though this does not necessarily mean that there is no way in which pirates can be influenced through the media. It is rather the method that was taken in the past, that simply did not resonate in the expected manner with the people that mattered.

2.3.1 Commercial Prevention



Figure 2: Anti-piracy promotional image from 1981

The commercial manner of trying to influence pirates to stop downloading has been a trajectory full of bumps. This is not in small part because the history of anti-piracy campaigns is one full of bad publicity and polarization (De Geest, 2012).

A perfect example of this is the campaign that the British Phonographic Industry (IBP) held during the 80’s, with the catchy slogan: “Home taping is killing music”². This campaign manifested itself due to fear that the turnover of albums would lower because technological innovations had made it possible to tape songs from the radio with the use

of cassettes. The logo and the slogan were the target of many comments and parodies. The band *Dead Kennedys* released their EP on cassette, on which they left a side blank and had it accompanied by the following message: “Home taping is killing record industry profit! We left this side blank so you can help” (Cummings, 2013). This campaign has since been handily used by the pro-piracy camp as a form to minimize the consequences of piracy and to additionally ridicule the opposition. With use of slogans like “Piracy is

² https://en.wikipedia.org/wiki/Home_Taping_Is_Killing_Music

theft” and the attitude that users of download methods are thieves, the music industry creates polarization and it shows that it does not learn from its mistakes.

In addition to the way that campaigning influences the opinions of piracy, there is also the case of the factual effects of downloading on revenue of the copyright holder and the way that these facts are presented. A critical look at the industry statistics of movie piracy calls into question the viability of the losses claimed by film studios. The MPAA (Motion Picture Association of America) claims that the major U.S motion picture studios lost \$6.1 billion in 2005 to piracy (L.E.K., 2005). These figures are based on the incorrect assumption that each item pirated equals a sale lost, but there is no guarantee that the pirate would have otherwise paid to see the film in a theatre (Yar, 2005). Lessig (2004) points out that in the same period that the RIAA (Recording Industry Association) estimated that 803 million CDs were sold, the RIAA also estimated that 2.1 billion CDs were downloaded for free. But, although 2.6 times the total number of CDs sold were downloaded for free, sales revenue fell by just 6.7 percent. This indicates that the ‘one CD downloaded is a sale lost’ idea is a fallacy. To punctuate this point, Lessig continues with:

"If every download were a lost sale then the industry would have suffered a 100 percent drop in sales last year, not a 7 percent drop. If 2.6 times the number of CDs sold were downloaded for free, and yet sales revenue dropped by just 6.7 percent, then there is a huge difference between downloading a song and stealing a CD."

Additionally, Yar points out that these statistics are generated by an industry with a vested interest in maximizing the figures and that there is no mechanism that prevents them from doing this. The more the government focuses on the issue, through heavier enforcement and litigation as well as tightening copyright laws, it produces new estimates suggesting that the 'epidemic' continues to grow unabated; which then legitimates industry calls for even more vigorous action. But what the industries seem to forget is the simple fact that it is common knowledge that the numbers that they present are not entirely factual. The way that the industries keep insisting on giving outrageous numbers as an indication of their losses – even though the proof of the inaccuracy of this is there for everyone to see – only fortifies the negative perception that the industries call upon themselves.

2.3.2 Media Influence and an Answer to Research Question 1

The purpose of this section was to determine the influence of social learning in having people commit piracy, and whether this influence was positive or negative. In contrast to the influence of outcome expectancies, it is difficult to determine the effects of the social learning factor as either being positive or negative. While day-to-day influence of close peers will obviously play an important part, it is challenging to measure the exact influence this has without focusing a research on it. Especially when considering that peer influence can differ immensely between individuals.

Our first research question was: which external factors influence people to commit piracy, and do they do this in a positive or negative way? We found that there are two external factors in particular that influence people in the attitude towards downloading. Outcome expectancies has a negative influence and therefore pushes people to download. Social learning is too different for individuals to determine positive or negative influence. Finding a social influence that is the same for everyone, however, could provide a measurable influence that determines if social learning is a positive or negative external factor. This measurable social influence is the media.

When looking at the history that media has played in the depiction of piracy, it was mostly influenced by companies protecting their copyright and how they decided to create the image of piracy in the media. That the negative depiction did not have the desired result is mostly a consequence of the way that the companies decided to implement it, resulting in neither a positive or negative social influence. But there is also a depiction of piracy that does not come from the companies that have a vested interest in painting it a certain way: from the independent media outlets. They also influence people by creating an image of piracy. And if several mass media outlets paint the same picture then it follows that this is copied by many of its viewers or readers (Bandura, 2002; Potter, 2012). Through a sentiment analysis of media depiction, I will therefore try to determine if media depicts a positive or negative image of piracy to see if this could be a positive or negative influence on someone's attitude towards piracy.

3. Sentiment Analysis: Method

Before we can truly start the research of the depiction of piracy in the media, we need to shed some light on the specifics of the research. The second research question - was the media depiction of intellectual property piracy in Dutch written media positive or negative between 2009 and 2014? - requires some illumination on the definition of certain terms. While words like ‘image’, ‘piracy’, and ‘media depiction’ seem straightforward, it is detrimental to the research that they are defined in a proper manner as to what they mean within the confines of this research. This ensures that there will be no doubt about what is meant when any of these words are later used within the research itself or within the analysis. This will be discussed in Section 3.1. In the following section (3.2), the exact method that was used for the sentiment analysis will be reviewed. This consists of an overview of how the corpus was defined as well as an explanation of the algorithm that was used to determine sentiment.

3.1 Defining and Explaining Key Components of Research Question 2

In this section, the following questions will be answered to illuminate all aspects of research question 2: what is media depiction; what is intellectual property piracy; and why this timeframe?

3.1.1 What is Media Depiction?

The question for this section is not necessarily what the definition in general of media depiction is, but rather how it is delineated within this research. In theory, any mention of illegal piracy within any media form is a media depiction of the topic, but it would be almost impossible to measure all the references and to draw relevant conclusions from the gathered data. Media depiction is therefore confined to one media form for the purposes of this research: the written media. There might be two questions in particular that arise with this decision: what exactly is meant by written media and why do we focus on the written media? By answering these questions, the definition of media depiction within this paper will be clarified. This will ensure that the term can be used freely without the need for continuous clarification.

1. What is the written media?

Before the other questions regarding media depiction can be answered, it first needs to be clear what exactly is meant with ‘the written media’. For the purposes of this paper, it does not include written blogs, or gossip magazines, but simply exists in two forms: as popular newspaper or news website. The exact line of demarcation that is used to determine if a newspaper or website is popular, is further explained later in this chapter, but it can be assumed that both forms have more than 50,000 readers per year.

2. Why the written media?

People gain news in all kinds of manners; it is no longer a case of a single news outlet providing the sole information on a subject. In addition to different news outlets providing information, social media attributes by having people bilaterally sharing important news. For this research, however, social media will not be measured as a source of information, and the focus will be on unilateral information provision, which is provided by news outlets. These companies can create an image of something which can then be taken over by the readers. When focusing on these media forms, there are two contenders for being the most influential: the written media and TV (Barry & Fulmer, 2004).

Though it is not impossible to determine sentiment on media focused on video (such as the daily newscast), it would require an additional step of transcribing the spoken text. The real difficulty in analysing this media form, however, lies in finding the relevant videos in the first place. While articles written in newspapers and on news websites are easy to filter on subject, it is harder to do this for a video. If the written media would be a worse representation of society than TV, it would be a reason to analyse TV broadcasts, but that is not the case (Wennekers & de Haan, 2017).

3.1.2 What is Intellectual Property Piracy?

Where the definition of media depiction required a clear delineation, the definition of intellectual property piracy is simpler. If someone’s intellectual property is downloaded without permission of the author, then it is piracy. Note that it doesn’t matter if you upload or download a work, both acts are piracy and currently they are also both illegal in the Netherlands. Reaching this point has been quite a journey, and it was therefore elaborated on in Chapter 2. The definition of the term does not change throughout the research, whenever downloading of someone else’s intellectual property is mentioned it is registered as piracy. Of course, it is almost impossible to analyse every mention of piracy in the written media, which is why my method includes a categorization of subjects that the research focuses on. This will be detailed later on in the chapter.

3.1.3 Why This Timeframe?

The timeframe specified in the research question might seem odd. Intellectual property piracy has been around as early as 1998, when Napster was released, and could therefore have been included in the research. The reasoning behind this is twofold: first, the timespan between 1998 and 2017 would have been too large to properly research all the relevant articles; and second, torrent website The Pirate Bay was blocked in 2009 from being visited by Dutch users. When this occurred, The Pirate Bay had been around since 2003 and had seen its number of unique visitors rapidly increase – it was one of the most popular websites for pirates. The block created waves of controversy from different sides, some argued that it impeded on the basic freedom of the internet and others said that it was a correct move to block a website that knowingly provided illegal files. In other words, many people had an outspoken opinion on the matter, which is the ideal time to measure sentiment. When other actions against piracy were later announced, the discussion would consistently flare up with the new information implemented. The timespan is closed at 2014, which is the year that The Pirate Bay came back online.

3.2 The Corpus

Determining what the sentiment of media outlets is towards a certain subject requires a fitting method to execute this. The research, in its basis, is simple: review as many of the written media as possible and determine their sentiment towards piracy - the data that results from this should give an answer as to whether the media depiction was positive or negative. Since it is impossible to review all that was ever written about piracy, it is important to create a corpus that can function as representative of the written media and the image they paint of piracy. In Section 3.2.1, the corpus will be created by determining the written media outlets and articles that will be analyzed. Section 3.2.2 is dedicated to detailing the algorithm that was used to rate sentiment in the articles.

3.2.1 Creating the Corpus

The selection of media outlets and articles was based on several experiences and ideas:

- While studying the different written media outlets in the Netherlands, it became clear that several newspapers (such as large newspapers De Volkskrant and Algemeen Dagblad) in the Netherlands have the same owner and therefore tend to recycle news articles from each other. To prevent the same news articles from being analyzed, only one of these outlets could be included within the corpus.

- The timeframe of the research question is between 2009 and 2014; if outlets were not present during that time, then they are not included in the corpus.
- Articles that only copy the information that was given by Dutch news agency ANP are not part of the research. The articles released by ANP are meant for all news outlets and therefore do not convey sentiment. In addition, the exclusion of these sorts of articles ensures that articles that are represented in the research are not almost exactly the same.
- News websites and newspapers are different forms of media and are therefore evenly represented to ensure that there is no bias towards one form.
- Not all written media outlets had articles that were easy to find on the internet. Articles from the newspapers were found using LexisNexis³ – outlets that were underrepresented on this database were not considered to be part of the corpus. Articles for news websites were taken from the websites themselves.

However, the amount of viable media outlets still seemed too large to investigate. To further narrow it down, the popularity of the media outlet was measured. Since the goal of the research was to measure influence of media on the opinion of the general populace, the inclusion of the most popular news outlets in the corpus was a logical action. For newspapers, this was done by looking at the number of newspapers that are in circulation⁴ and for news websites this was measured by the cost of an advertisement that is featured with an article when it is posted on the website (Smeenk, 2016). The distinction between the measurements of the popularity is twofold: First, the circulation of newspapers is public knowledge, while websites often do not publish the number of unique visitors that they have per day. As a consequence of this information not being public, the only real estimate about the amount of people that will read an article on a website can be deduced from what the website thinks its advertisements are worth. If thousands of people read an article, an advertisement will be more expensive than when only hundreds of people read it.

De Telegraaf (circulation of ~455.000) was the most popular newspaper and Nu.nl (885 euro for an advertisement) was the most popular news website. For news websites, the next on the list was easy to determine; the two other most popular websites were ones that are directly affiliated with newspapers and therefore of no interest to this research (as websites that function as extension of a newspaper could probably be its own category) – Rtlnieuws.nl became the next option for an analysis. For newspapers, it was Algemeen Dagblad that had the next highest circulation, but relatively few articles were available on LexisNexis and around a quarter of these articles were copied from ANP. Consequently, I chose to analyze articles from Trouw, which has the same owner as Algemeen Dagblad and did have more articles available on LexisNexis that discussed piracy. With the mainstream news providers covered, it seemed imperative for the third choice of website and newspaper to be slightly different to ensure that all layers of society are represented in the

³ A Dutch news article database (<http://www.lexisnexis.com>)

⁴“National Onderzoek Multimedia” (<http://www.nommedia.nl/1854400/Print-oplage.html>)

research. For websites, once again, the choice was easily made – Tweakers.nl is enormously popular but focuses mostly on subjects related to IT, which results in a multitude of articles discussing all forms of piracy. For newspapers, taking a local newspaper could provide a different insight into how piracy is depicted, especially since local newspaper mostly have an older audience (Mitchell, 2016). The only newspaper that fit the bill and had enough articles about piracy was Het Parool. The complete list is:

Newspapers	News websites
De Telegraaf	Nu.nl
Trouw	Rtlnieuws.nl
Het Parool	Tweakers.nl

Figure 3: Media outlets that will be analyzed on sentiment

With the media outlets chosen, the articles needed to be filtered. The most important reason for doing this is to narrow down the corpus to a reasonable size, but an additional reason is to have articles that can be easily compared. If articles cannot be properly compared, we can never arrive at the conclusions that one media outlet is more prone to negativity or positivity than another. The sentiment of the individual outlets is not immediately relevant to the research, but if a negative or positive peak in sentiment occurs, it will have to be determined how it happened. When the articles cover the exact same subjects, determining what exactly the cause is becomes easier. The reason could range from it being a biased outlet, a biased writer, or simply a tonal difference.

The articles themselves will be filtered based on historical events of piracy to ensure that the articles are referring to similar events. The articles will have to be written between 2009 and 2014 and most are focused on torrent website The Pirate Bay, as this has been a major discussion point in news coverage. The events are:

The Pirate Bay

1. The Pirate Bay is blocked in the Netherlands (2009).
2. The Pirate Bay is incriminated by Brein; hackers put down Brein’s website (2009).
3. Brein asks for a blockade of The Pirate Bay – Internet Service Providers refuse (2010).
4. The judge determines that The Pirate Bay will have to be blocked (2012).
5. Alternatives for The Pirate Bay arrive (2012).
6. Blockade of The Pirate Bay is removed as the judge determined it to be of little use (2014).

Download ban

1. European Court decides that the Netherlands will have to implement a download ban (2014).

2. The Netherlands gets a download ban – what now? (2014)

MegaUpload

1. FBI raids MegaUpload offices, which results in the website being taken down (2012).

SOPA (Stop Online Piracy Act)

1. Big websites, like Wikipedia, go ‘black’ in protest as answer to the anti-piracy law SOPA (2012).

With this division, all media outlets had one article that fit the bill for the subject. Some outlets cover subjects that slightly differ from the one mentioned above, but that still touch upon that same subject. If such an article is used in the research, it is because no other articles of that outlet would come closer to covering that subject.

3.2.2 The Algorithm

Performing a sentiment analysis requires some specific rules to ensure consistent analysis. When interpreting 60 articles of varying length based on their sentiment, consistency is a detrimental factor. This consistency could be achieved by assigning value to singular words, such as always evaluating ‘stealing’ as a negative word. But this creates a pretty flat sentiment analysis in which the tone of the author could be missed. Another way, and the way in which the analysis will be executed for this research, is to create general rules that do not try to assign value to singular words but rather to the words and the sentence, or context, that it is contained within (Godbole, 2007). There are several things to consider such as the depth of a sentence, or the fact that two words that are close together could have opposite polarity when they are separated by “and” from when they are separated by “but” (Hatzivassiloglou, 1997). By predetermining the sentiment for when this happens, through the use of an algorithm, a consistent analysis should arise.

The algorithm will help determine whether an article is positive or negative towards piracy - this will consist of a set of rules that state the points given when something occurs. Since this research focuses on sentiment towards a single topic, this is somewhat differently done in comparison with most other forms of sentiment analysis. Just calculating if an article is positive or negative is not sufficient for the case under investigation, and words are therefore not graded as either positive or negative, but rather as ‘positive towards piracy’ or ‘negative towards piracy’. Though the distinction may seem small, this relies more heavily on the context of the word and therefore requires a clearer ruleset to make it easier to understand why something was rated in the way that it was.

The algorithm consists of 15 steps (with the letter being a short explanation of or an expansion upon the rule):

1. A noun or verb that is negative towards piracy will add -1 point to the total sentiment of an article; a noun or verb that is positive towards piracy will add +1 point to the total sentiment of an article.
 - a. The sentiment of singular words often speaks for itself, but sometimes requires a clarification. This can be found in the comments that accompany the article (Appendix A).
2. A negative adjective will add -1.5 points to the total sentiment of an article if it is followed by a negative noun; a positive adjective will add +1.5 points to the total sentiment of an article if it is followed by a positive noun.
 - a. The 1.5 point rating is determined to be as such as it was felt that adjectives in connotation with nouns increase the sentiment beyond that of the individual words. The noun themselves are also still graded, giving a sentence like 'the ugly lie' (0 -1.5 -1) a -2.5 sentiment.
3. A negative adjective that is followed by a neutral noun will add -1 point to the total sentiment of an article; a positive adjective that is followed by a neutral noun will add +1 point to the total sentiment of an article.
4. A positive adjective that is followed by a negative noun will add -0.5 points to the total sentiment of an article; a negative adjective that is followed by a positive noun will add +0.5 points to the total sentiment of an article.
 - a. The best example of this would be "The beautiful lie" which is classified as /determiner adjective adverb/ and would normally get +0 points (0 +1 -1). However, this does not seem to convey the true sentiment of the sentence – which should be negative as, though it may be beautiful, it still concerns a lie. Nouns and adjectives are therefore seen in pairs, if both are positive or negative the other rules apply, but if they are of opposing sentiment, they are graded together. "The beautiful lie" therefore receives a -0.5 as sentiment (0 ~~(+1-1)~~ -0.5).
5. A neutral adverb that enhances the sentiment of an adjective that is negative towards piracy, will add -0.5 point to the total sentiment of an article; a neutral adverb that enhances the sentiment of an adjective that is positive towards piracy will add +0.5 point to the total sentiment of an article.
 - a. Depth of a sentence will now play a part as well, as is the case with: "The very beautiful lie". The noun and adjective are graded together in accordance with rule 4 – should the adverb influence them as a set? For the purpose of this research, it will not. Meaning that the adverb only receives a grade that comes from the adjective. "The very beautiful lie" is therefore graded as +0.5 (0 +0.5 ~~(+1+1)~~ -0.5).
6. A negative adverb that is followed by a positive or negative word or expression will alter the sentiment of that word or expression to negative. A positive adverb that is followed by a negative word or expression will alter the sentiment of that word or expression to positive.
 - a. An example of this would be: "Not lying". The verb would normally be graded with -1 sentiment, but the existence of the 'not' changes the sentiment of the verb to +1; as the exact opposite of a negative thing is happening.

7. A negative adverb that is followed by a neutral word will add -1 point to the total sentiment of an article; a positive adverb that is followed by a neutral word will add +1 point to the total sentiment of an article.
8. If someone negative towards piracy is quoted, it will add -2 points to the total sentiment of an article; if someone positive towards piracy is quoted it will add +2 points to the total sentiment of an article.
 - a. Quotations are an important part of news articles, and while reading the news items that were going to be analyzed, it showed that articles have the tendency to only let one party speak. This is, of course, a bias towards one party and it should therefore be reflected in the sentiment analysis.
 - b. A quotation itself is rated in accordance with rules 1 through 7.
 - c. A quotation will only be classified as such when someone is directly quoted.
 - d. If someone is quoted multiple times throughout the article, the sentiment points will only be given once and the following quotations will be analyzed as the rest of the text.
9. Expressions that are negative towards piracy will add -1 point to the total sentiment of an article; expressions that are positive towards piracy will add +1 point to the total sentiment of an article.
 - a. Expressions are rated in their entirety and the sentiment of individual words are therefore not counted. 'Having all his ducks in a row' is, for instance, an expression that contains neutral words, but which expresses a positive sentiment and will therefore add +1 point to the total sentiment of an article.
10. When an article normalizes piracy it will add +2 points to the total sentiment of an article.
 - a. An example of normalizing piracy would be an article which speaks directly to the reader as a downloader. "You do not have to worry about getting caught even though a download ban was implemented" does not use positive or negative words, but clearly displays its feelings towards piracy. Because 'normalization' is a broad term, every time this rule is in effect, an explanation for it can be found in the comments that accompany the article (Appendix A). There are 2 points assigned to this kind of sentiment because of the impact it can have on a reader as it can actively tell them that there is nothing wrong with piracy.
11. When a negative aspect of piracy is mentioned which does not use negative words towards piracy, the entire section that describes this aspect will add -1 point to the total sentiment of the article; when a positive aspect of piracy is mentioned which does not use positive words towards piracy, the entire section that describes this aspect will add +1 point to the total sentiment of the article.
12. Negativity or positivity towards the 'homecopy tax' will be graded as neutral towards piracy.
 - a. Though the tax is a consequence of piracy, there seems to be a general consensus that it is inefficient and that many people are not even aware of its existence. Being against the tax therefore does not necessarily indicate a certain feeling towards piracy – it just demonstrates

its ineffectiveness. A positive statement towards the tax merely shows that artists lose out on income because of piracy, which certainly cannot be denied.

13. Every repetition of a word is counted towards the sentiment of an article.
14. Negativity towards the exuberant fines that certain countries give out to people who download copyrighted material, will be graded as positive towards piracy.
 - a. An argument towards not giving out high fines can be made by both camps in the piracy-debate. Yet, when an article specifically mentions that they are against it, they are implying that piracy is actually not similar enough to a crime to be fined in that manner. This contributes to the normalization of piracy, which is a rule in this research that constitutes a statement to be positive towards piracy.
15. Mentioning the ‘freedom of the internet’ as an aspect of piracy will be graded as positive towards piracy.
 - a. Many actions that are taken against piracy, actively put the freedom of the internet in danger and most articles argue that this should not be the case. But, by doing so, they prioritize piracy above the basic ability of the industry to protect their copyrighted material.

The implementation of the algorithm is done by hand. This means that all 60 articles were analyzed by me, through the use of the ruleset depicted in the algorithm. When in doubt about the sentiment of a word or sentence, the public lexicon SentiWordNet⁵ was used as an aid. This lexicon offers a sentiment when inserting a string of text. All articles and their rated sentiment are available as an appendix to this paper.

⁵ <http://sentiwordnet.isti.cnr.it/>

4. Sentiment Analysis: Results

Our theoretical frame has now put the situation in perspective. The method has been explained in the previous section, so it is time to perform the sentiment analysis in order to find out whether the media portrayed piracy in a positive manner. The results from our sentiment analysis are given per historical event as they are listed in the previous chapter. For every event, some observations are made about the dataset that emerged to give an indication as to what created those results and what is noticeable about it. After all the data is presented, conclusions are derived from the resulting data. Before we get into the results of the research, there is one special case that needs to be highlighted, as it drastically influenced the results.

4.1 The 'Illegal' Issue

Throughout the course of analyzing the articles, there was one issue that kept appearing: the word 'illegal' showed up an exponential amount of times in comparison with other non-neutral words. Though it seemed simple to classify the word in all its variations as negative, the degree in which it appeared at least asked for a consideration. Throughout the articles it was used 125 times – which was 63 times more than the second most graded word. It influenced almost all articles in their sentiment and it completely changed the sentiment of an article from positive to negative in 8 cases. In all these cases 'illegal' was graded as negative sentiment, as it is clearly a negative word – especially when taking into account that piracy was not necessarily illegal when most of the articles were written. But with the impact that this word had, a new look was required and it turned out that the word often was not negative at all. 71 mentions of 'illegal' were found in articles that discussed the download ban. Many of these articles have at least one appearance of the word when they quote the judge in its verdict or quote Brein in reference to the verdict. In addition, it is often used in the context of 'the war against illegal downloading' which is an expression. Since the word 'illegal' was therefore not necessarily chosen by the writer itself, it does not reflect negative sentiment. With this in mind, a change to the rules was easily made: whenever 'illegal' was used in a quotation of the verdict or as an expression, its sentiment was not counted. This created a more consistent dataset, where the word 'illegal' no longer changed the sentiment of articles or where the word was graded way more often than others.

4.2 Results

Below, the results of each sentiment analysis per historical subject (Download ban, MegaUpload, SOPA, The Pirate Bay) are presented. Each segment of the results starts with a short introduction surrounding the subject,

followed by the events that the articles cover, a graph of the results, and some observations about the outcome of the analysis.

Download ban

The download ban was introduced in the Netherlands in 2014 and caused quite some controversy. Where it was not illegal to download before this new law, as long as the file was not shared with others, this changed with the new ban. All copyrighted material was now protected and was not allowed to be downloaded at all. The events that the articles cover are:

Article 1: European Court decides that the Netherlands will have to implement a download ban (2014).

Article 2: The Netherlands gets a download ban – what now? (2014)

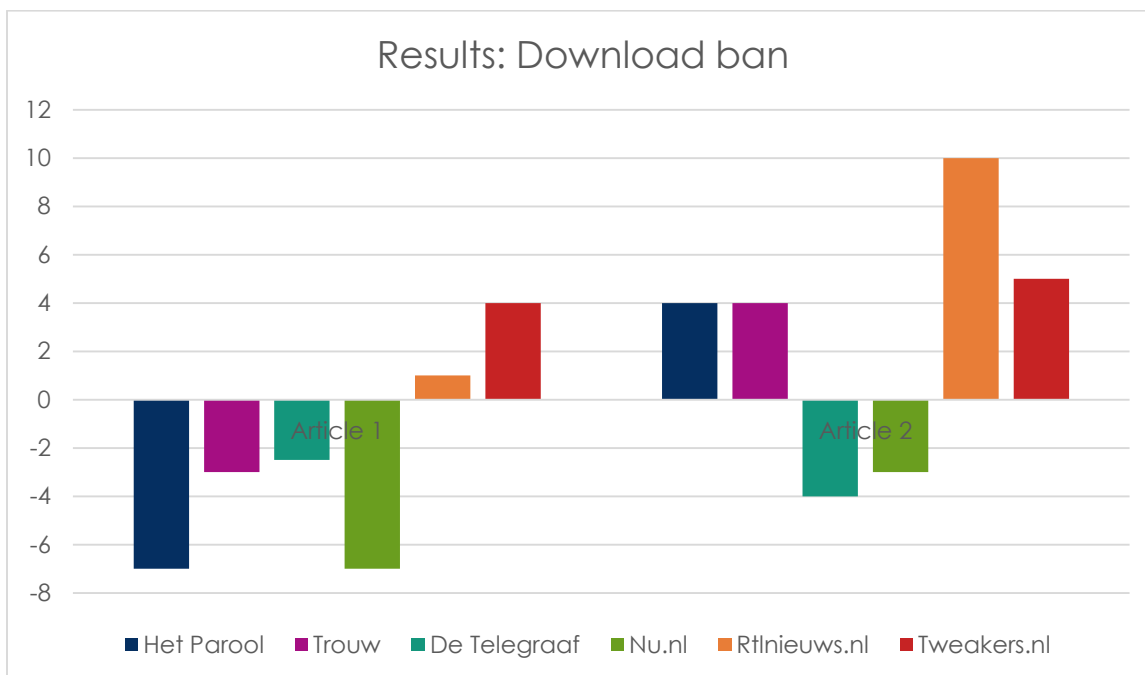


Figure 4: Results from articles about the download ban

The sentiment in the articles concerning the implementation of the download ban are almost all negative towards piracy. Interestingly enough, most of them do not necessarily concern themselves with piracy itself but rather take a negative approach towards the homecopy tax and a rather positive approach to the download ban. It is this positivity towards the ban that causes the results to be negative.

The second article is clearly more slanted towards positivity, as the subject lends itself to normalization of piracy which is graded as positive towards piracy in rule 10 of the algorithm. But the inclusion of the

subject is based on the fact that every media outlet wrote an article about it, which might indicate more about the stance towards piracy of the outlets than it does about the subject itself.

MegaUpload

MegaUpload was a website that used the direct-download model to share files with others. In 2012 this website was taken down on a court order of the United States Department of Justice. The articles used for this subject cover that event.

Article 1: FBI raids MegaUpload offices, which results in the website being taken down (2012).

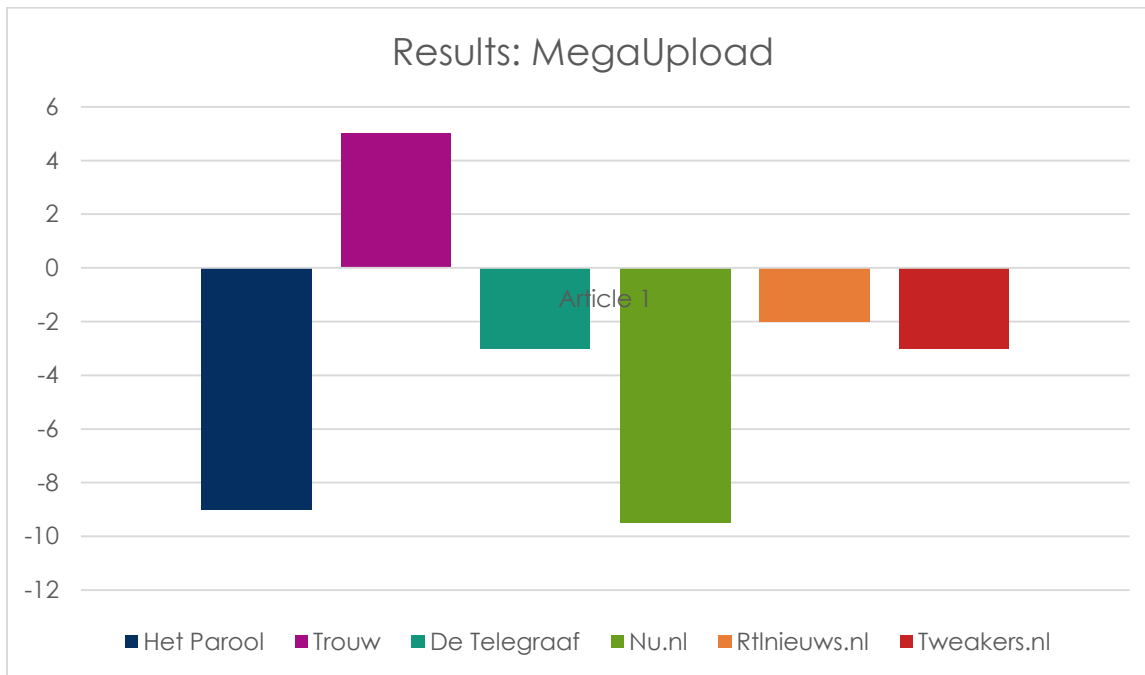


Figure 5: Results from articles about MegaUpload

The primarily negative results for these articles are no real surprise as they discuss the downfall of one of the bigger download sites. In addition, hacker group Anonymous started interfering which lend itself for some negative comments in the articles.

Of particular note on the MegaUpload topic is that when downloading is called ‘illegal’ that is incorrect information. Using The Pirate Bay always meant that a downloader was also uploading – which has been illegal for a long time. However, in 2012, just downloading was not illegal if the file was not shared with anyone else. This gives an indication of the negativity surrounding this topic.

SOPA

The ‘Stop Online Piracy Act’ was a controversial American bill that expanded the ability of United States law enforcement on the internet to prevent piracy. Since this heavily impacted the privacy of all internet users, many people, and companies, were against the bill. This is displayed in the subject of the articles.

Article 1: Big websites, like Wikipedia, go ‘black’ in protest as answer to the anti-piracy law SOPA (2012).

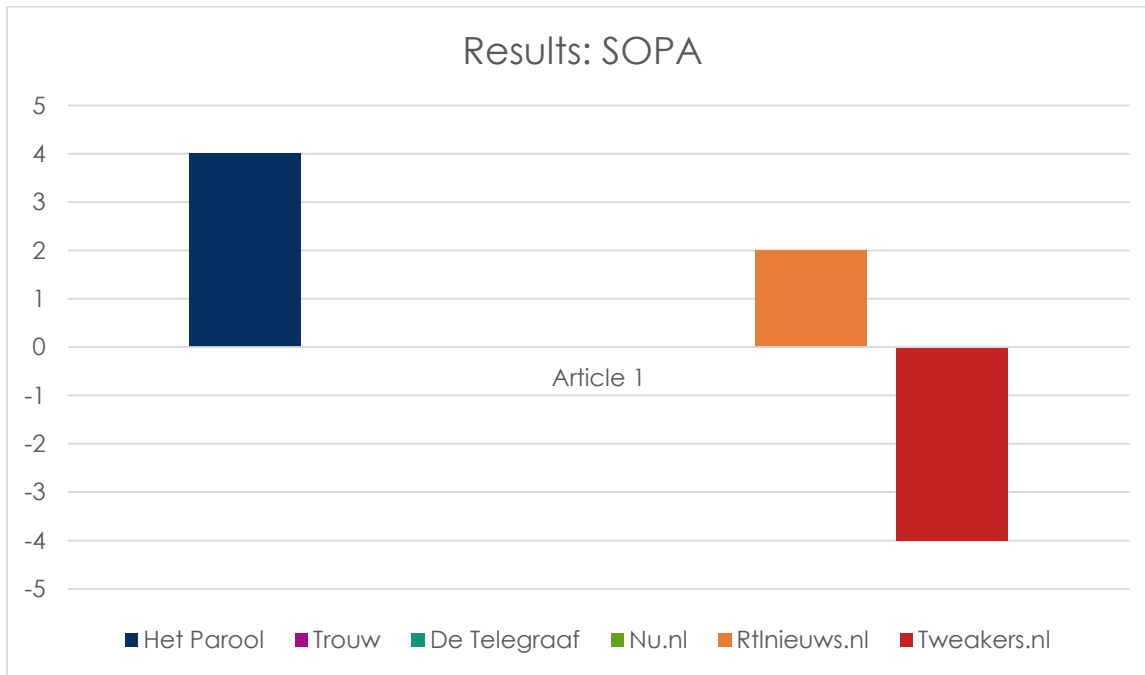


Figure 6: Results from articles about SOPA

A large reason for the inclusion of this topic in the research, was because it lends itself for a discussion on piracy. Here is a way to actively fight piracy, but could the consequences that it carries be worth it? However, it turns out that the articles did not bring this up in any real substantial way and piracy is hardly mentioned at all. Therefore, the articles are mostly neutral, except Het Parool which is an advocate of the free internet (and Tweakers simply uses a few negative words but the article does not read as really negative). Nu.nl, Trouw, and De Telegraaf all had a sentiment of 0 and are therefore not seen in the graph.

The Pirate Bay

The Pirate Bay is one of the biggest websites to offer links to copyrighted material that is shared by other users. Because of the enormous amount of available material on The Pirate Bay, it has been the most targeted website in the war against piracy. The consequence of this is that many events have occurred over the last few years surrounding this particular subject. The articles cover several of these events.

Article 1: The Pirate Bay is blocked in the Netherlands (2009).

Article 2: The Pirate Bay is incriminated by Brein; hackers put down Brein’s website (2009).

Article 3: Brein asks for a blockade of The Pirate Bay – Internet Service Providers refuse (2010).

Article 4: The judge determines that The Pirate Bay will have to be blocked (2012).

Article 5: Alternatives for The Pirate Bay arrive (2012).

Article 6: Blockade of The Pirate Bay is removed as the judge determined it to be of little use (2014).

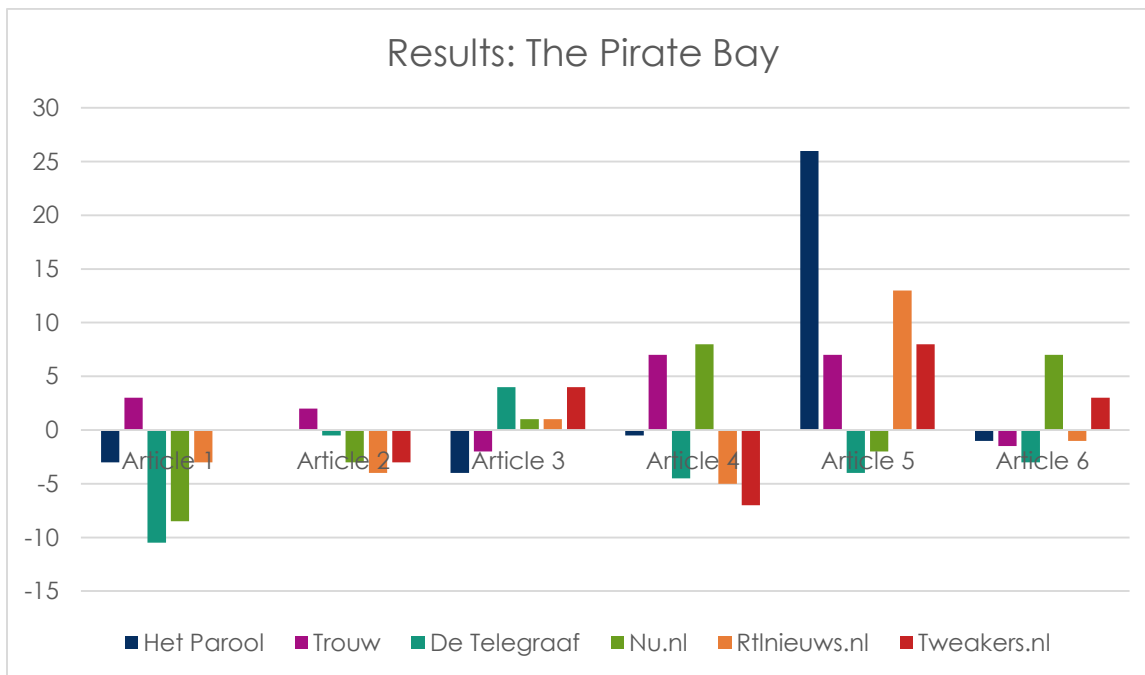


Figure 7: Results from articles about The Pirate Bay

Arguably the heart of the research, acknowledged by the number of articles handled, this topic has an interesting scale that increases over time. Starting out mostly negative, the last few articles are graded as mostly positive towards piracy. However, keeping in mind the dates of all the articles, it means that the SOPA and MegaUpload articles were written after article 5, and the articles concerning the download ban were written after article 6. This means that the sentiment increase is not as linear as it would seem in the graph above.

4.3 Overview of the Data: an Answer to RQ2

With the sentiment analysis executed, it is time to make a summary of the results. This data will first be presented as it emerged from the research, after which some small tweaks will be made to make it more accurate. Afterwards, a short conclusion will be drawn from the results.

In total, there were 60 articles analyzed from 6 different media outlets; 51.6% of these articles gave a negative result, 38.4% was positive, and 10.0% was neutral. The results per outlet are depicted in table 3. Het Parool, for instance, had 5 negative articles, 4 positive articles, and 1 neutral article:

	<i>Het Parool</i>	<i>Trouw</i>	<i>Telegraaf</i>	<i>Nu.nl</i>	<i>Rtlnieuws.nl</i>	<i>Tweakers.nl</i>	Total
<i>Negative</i>	6	3	8	6	5	4	32 (53.3%)
<i>Positive</i>	3	6	1	3	5	5	23 (38.3%)
<i>Neutral</i>	1	1	1	1	0	1	5 (8.3%)

Table 1: Sentiment of articles

This data suggests that negativity towards piracy is more prevalent than positivity. However, a positive or negative value could be anything that is not 0, so the question then remains as to how those sentiments were distributed. For instance, Het Parool had six articles with negative sentiment of which the sentiment ratings were: -7, -9, -3, -4, -0.5, and -1. Together, these create a -24.5 negative sentiment for Het Parool. The total points for the other media outlets can be found in the table below:

	<i>Het Parool</i>	<i>Trouw</i>	<i>Telegraaf</i>	<i>Nu.nl</i>	<i>Rtlnieuws.nl</i>	<i>Tweakers.nl</i>	Total
<i>Negative</i>	-24.5	-5.5	-32	-33	-15	-17	-127
<i>Positive</i>	34	28	4	16	27	24	133

Table 2: Total sentiment points

From this data, it becomes clear that negative sentiment in an article is less weighted than positive sentiment is. In other words, if an article is positive towards piracy, it uses more positive sentiment than it uses negative sentiment for a negative article. Therefore, even though the total amount of articles with a negative sentiment is higher than articles with a positive sentiment it does not necessarily mean that negative sentiment is most projected on the readers. To further distill the data from this point of view, a look at the articles with small sentiment is also necessary. If an article has a sentiment of -1, it is difficult to say that it has a negative sentiment – on paper it certainly is, but in reality these articles often turn out to be neutral with just one or two

words to change sentiment. By classifying these articles as neutral, instead of positive or negative, a more accurate dataset emerges. The articles with a sentiment below 3 will therefore be classified as neutral. This means that a sentiment rating of +3 will indicate a positive sentiment and +2 will indicate a neutral article. This particular line of demarcation arises from the results of the research itself; with a sentiment of 3 or more there are often more rules that are invoked to reach that number. Of the five times that a sentiment rating of 2 was given, three of these were reached with the use of only two rules.

The results from this modification are:

	<i>Het Parool</i>	<i>Trouw</i>	<i>Telegraaf</i>	<i>Nu.nl</i>	<i>Rtlnieuws.nl</i>	<i>Tweakers.nl</i>	Total
<i>Negative</i>	4	1	6	5	3	4	23 (38.3%)
<i>Positive</i>	3	5	1	2	2	5	18 (30.0%)
<i>Neutral</i>	3	4	3	3	5	1	19 (31.6%)

Table 3: Total sentiment of articles with small results as neutral

From this data, we can conclude that negative sentiment slightly outweighs positive sentiment but that neutrality and positivity are also heavily conveyed in the articles. The overall sentiment is therefore neutral.

All in all, the answer to research question 2 is that Dutch written media portrayed piracy in a mostly neutral manner between 2009 and 2014. This means that media does not actively try to deter people from downloading. In the following chapter, the impact of this will be determined and some possibilities for additional research are posed.

5. Conclusion & Discussion

In this last chapter, a summary of the results of the research will be given first (Section 5.1) to give an overview of all that was discussed. In Section 5.2, the context of these results will be given to create an environment that makes it possible to answer the problem statement as it was posed in the Introduction. In section 5.3, possibilities for further research are discussed.

5.1. Summary of the Results

The aim of this research was to determine why many people download without remorse even though they know that it is considered to be stealing. The goal was to possibly find an additional method of combatting piracy. To examine this, the first focus was on what external factors influence people to commit piracy and whether they do this in a positive or negative way. Through the use of Lowry's SCT framework, which was derived from a meta-analysis based on piracy-related research, two influential factors were found. Of these two factors, it was the outcome expectancies of piracy that was the clearest factor to push people towards illegal downloading. The reasons for downloading are clear for anyone who might want to download, and the reasons to not download are scarce. The Dutch law has been changed over the last few years to ensure that negative outcome expectancies emerge, but the changes have not been substantial enough to deter the pirates from downloading. Social learning, the other external factor that was investigated, offers less clear results. While it is obvious that peers influence someone in their attitude towards illegal downloading, it is a bit harder to determine exactly how they do this and in what capacity. By focusing only on the influence of media on the behavior, this research aimed to highlight the most consistent and most measurable peer. A sentiment analysis followed from which is concluded that the media is a little negative towards piracy but is almost equally positive and neutral. This neutrality shows that media does not deter pirates from downloading. The combination of external factors and social learning then tips the scale for pirates towards downloading instead of against it. Outcome expectancies have been a point of scrutiny, but it has been hard to alter the influence of this factor. Knowing that media depiction of piracy is neutral shows a different influential aspect that has hardly been addressed and might prove to be a possible way to battle piracy.

The conclusions from this research only carry weight when put into perspective with their context. This context is therefore given below, and the consequences of the results are also discussed.

5.2. Context of the Results and an Answer to the Problem Statement

It will come as no surprise that research surrounding the influence of peers on digital piracy has been conducted (Lowry et al., 2017). However, much of the research has been focused on the influence of peers on criminal behavior as was posed by the Social Learning Theory. What these researches miss out on is the fact that many pirates do not see their actions as criminal behavior, which separates it from other criminal actions like 'regular' stealing. Additionally, many researches have been focused on the influence of mass media, but not necessarily on the attitude it displays towards piracy. This paper tried to find a middle way between these researches by focusing on the attitude towards piracy from both peers and the media. One final aspect to keep in mind surrounding this research is the fact that the internet is changing and becoming more open. Open Source and Creative Commons are new methods of removing the old forms of copyright where a singular person or company owned all that was associated with a work (Stallman, 2010). These new forms take a different approach where sharing is often encouraged instead of shunned. Creative Commons, for instance, provides simple, general, and non-exclusive licenses. By choosing for such a license, the right is made public for everyone to use the work for free and to reproduce and spread the work, for as long as they want. The only requirement is that the Creative Commons licenses (or a link referring to it) are included when a copy of the work is spread. In addition, there are many requirements and restrictions that the license can enforce: usage for commercial ends can be prohibited; noting the name of the author can be forced; as well ensuring that any derived works are given the same license. The Creative Commons are used by websites like Flickr and Wikipedia which consequently means that more than 100 million photos are under the CC-license. It is therefore important to keep an open mind when it comes to the way that we view piracy. With the way that copyright is currently structured, there is no doubt that piracy is illegal, but it is possible that the general opinion of this might change. If this happens, then the research might be used to ultimately show a difference in media portrayal before and after the changes.

With the context in mind, we can find a possible answer to the problem statement, which is as follows: Intellectual property piracy is illegal and costs the relevant industries significant amounts of money, yet many people download without remorse. Now that we know that media is not actively deterring people from downloading, it becomes clear that the image of piracy in the media needs to be changed. Piracy is illegal since 2014 and this neutral attitude therefore needs to change in order to convey to readers that it is no longer allowed. SLT shows that people are heavily influenced by peers, including the media, and perhaps a change in tone of such a peer might prove to be profitable in the war against piracy.

5.3 Further Research

Fully researching all external factors from SCT that influence someone towards downloading – outcome expectancies, self-efficacy and self-regulation, moral disengagement, social learning, and environmental and other factors – will offer better insight into all reasoning for a pirate. When all factors are taken into consideration, it will be easier to determine the value of the influence of the media when it comes to piracy behavior.

The use of sentiment analysis to determine the media depiction of piracy might benefit from scrutiny. With a corpus of 60 articles, any real conclusions about media depiction need to be limited to that corpus and statements about the entirety of media depiction are therefore hard to make. The use of a computer-based algorithm might prove to be able to handle a much larger corpus, which could give a better indication of media depiction towards piracy. If this teaches us that the influence of media towards digital piracy absolutely has been positive (that it moves people to download), then the consequences should be investigated. A possible influence could have been the changes in media – where more companies are moving towards internet which means that there are more choices for people from where they want to consume their media. If media websites have to appeal more to their audience in order to gain enough readers, then it might do them a disservice to be negative towards actions that their readers are taking.

Lastly, ways to change the attitude of media outlets need to be researched before it can be used as a means of battling piracy. It is possible that media outlets are currently not aware of their portrayal of piracy and could be persuaded to change it. Otherwise, different methods of changing attitude might need to be researched.

6. References

17 U.S. Code § 504 (1946, July 5).

ACI Adam BV and Others v Stichting de ThuisKopie, C-435/12, 2014.

Anderson, N. “The YouTube Effect: HTTP traffic now eclipses P2P”, *Ars Technica*, 2007.

Aleassa, H. “Investigating Consumers' Software Piracy Using an Extended Theory of Reasoned Action”, Southern Illinois University, 2009.

Bandura, A. *Social Learning Theory*. Oxford, England: Prentice-Hall, 1977.

Bandura, A. *Social foundations of thought and action*. Englewood Cliffs, NJ: Prentice-Hall, 1986.

Bandura, A., Bryant, J. “Social cognitive theory of mass communication”, *Media Effects: Advances in Theory and Research*, 2002.

Barry, B., Fulmer, I. “The Medium and the Message: The Adaptive Use of Communication Media in Dyadic Influence.” *Academy of Management Review*, vol. 29, no. 2, 2004.

Yung, C. "Hong Kong Freezes Megaupload Assets". *The Wall Street Journal*, 2012.

Choi, B.H. “The Grokster Dead-End”, *Harvard Journal of Law and Technology*, vol. 19, 2006.

Cummings, A.S. “Democracy of Sound: Music Piracy and the Remaking of American Copyright in the Twentieth Century”, Oxford University Press, 2013.

Depoorter B., Vanneste, S. “Norms and Enforcement: the Case against Copyright Litigation”, *Oregon Law Review*, vol. 84, 2006.

Elkin-Koren, N. “Making Technology Visible: Liability for Internet Service Providers for Peer-to-Peer Traffic”, *New York University Journal of Legislation and Public Policy*, 2006.

- Freeman, R.E., Hodrick, L.S., Parmar, B., Tuslow, W. "File Sharing and Napster", University of Virginia, 2008.
- Geest, J. "Technologie En De Afdwinging Van Het Auteursrecht in De Muziekindustrie", Universiteit Gent, 2012.
- Glanz, K., Rimer, B. K., & Viswanath, K. *Health behavior and health education: Theory, research, and practice*. Hoboken, NJ: John Wiley & Sons, 2008.
- Godbole, N., Srinivasaiah, M., Skiena, S. "Large-Scale Sentiment Analysis for News and Blogs", University of Vermont, 2007.
- Hatzivassiloglou, V., McKeown, K.R. "Predicting the Semantic Orientation of Adjectives", Proc. 8th Conf. on European chapter of the Association for Computational Linguistics, Association for Computational Linguistics, 1997.
- Higgins, G., Fell, B., Wilson, A. "Low Self-control and Social Learning in Understanding Students' Intentions to Pirate Movies in the United States", *Social Science Computer Review*, 2007.
- Hinduja, S. "Music Piracy and Crime Theory", New York: LFB Scholarly Publishing LLC, 2006.
- Honigsberg, P.J. "The Evolution and Revolution of Napster", *University of San Francisco Law Review*, vol. 36, 2002.
- L.E.K. "The Cost of Movie Piracy", 2005.
- Lessig, L. *Free Culture: The Nature and Future of Creativity*. New York: Penguin. 2004.
- Limayem, M., Khalifa, M., Chin, W. "Factors Motivating Software Piracy: A Longitudinal Study", *IEEE Transactions on Engineering Management*, vol. 51, 2004.
- Lowry, P., et al. "Nature or Nurture? A Meta-Analysis of the Factors That Maximize the Prediction of Digital Piracy by Using Social Cognitive Theory as a Framework." *Computers in Human Behavior*, vol. 68, 2017.

Metallica v. Napster, Inc., SCC 2206, 2001.

Mitchell, A. "Younger Adults More Likely than Their Elders to Prefer Reading News." Pew Research Center, 2016.

Poort, J., Leenheer J. "Filesharing 2012: Downloaden in Nederland", CentERdata (IVIR), 2012.

Potter, W. "Media Effects". Thousand Oaks, CA: SAGE Publications, 2012.

S.968.R.S. Protec IP Act, 112th Congress, 2011-2012.

Samuelson, P., Wheatland, T. "Statutory Damages in Copyright Law: A Remedy in Need of Reform", 51 Wm. & Mary L. Rev., 2009.

Smeenk, D. "Nu.nl Bovenaan Top 5 Meest Waardevolle Nieuwssites." Nederlands MediaNieuws, 2016.

Stallman, R.M., "Free Software, Free Society: Selected Essays of Richard M. Stallman", Free Software Foundation, 2010.

Steinmetz, K. "The Pixelated Jolly Roger: An Ethnographic Analysis of Online Pirates", Eastern Kentucky University, 2010.

Wenneker, A, de Haan, J. "Nederlanders en Nieuws, Gebruik van Nieuwsmedia via de Oude en Nieuwe Kanalen.", 2017.

Wet AW 1912. (1912, September 23). Consulted on 14 October, 2016.

Yar, M. "The Global 'Epidemic' of Movie 'Piracy': Crime-wave or Social Construction?", Media, Culture & Society, vol. 2, 2005.