UNIVERSITY OF GOTHENBURG

DEPARTMENT OF
APPLIED IT

PREDICTION ACCURACY AND AUTONOMY

Assessing how recommender systems objectives can align with user autonomy

Anton Angwald

Kalle Areschoug

Thesis:	15 hp
Program:	Cognitive Science
Level:	First Cycle
Year:	2021
Supervisor:	Alan Said and Alexander Almér
Examiner:	Andreas Chatzopoulos
Report nr:	0069

Abstract

Entertainment recommender systems have been criticised by journalists and tech-industry insiders for undermining individuals' autonomy. These systems might exercise unwanted control over peoples' lives, not through coercion but rather through distraction. In this thesis we adopt an interdisciplinary framework to explore how the design of recommendation systems for entertainment services can align with the individual right to autonomy.

First, we assess design objectives by doing a corpus analysis on 1,883 scientific articles on entertainment recommender systems. We then carry out a qualitative survey of psychological literature and connect findings on self-regulation, sense of agency and habits to the autonomy of users. We also survey relevant literature on user-centred interaction design to relate the notion of user autonomy with user value. Finally, we focus on the specific use-case of YouTube's recommender system and propose design changes aimed at better aligning service provider objectives with users' objectives.

We conclude that because of an intention-behaviour gap, users' behaviour is an inaccurate reflection of users' intentions. Because of this, only analysing behavioural data undermines users' autonomy. Many current recommender systems, including YouTube's, use behavioural data since the data is easy to collect and often maximise service providers' goals. We propose both corrective and preventive solutions to this problem. The corrective solutions focus on offering users more customisability. The preventive solutions focus on ways to gather more data that better correspond to users' intentions. Higher user customisability can provide user data that can be expected to correspond relatively well to users' intention.

Keywords

recommender systems, autonomy, design ethics, user studies, evaluation metrics, YouTube, intention-behaviour gap

Prediktion och Autonomi:

Hur rekommendationssystem kan designas för att förbättra användarens autonomi

Sammanfattning

IT-industrin har kritiserats för att utforma applikationer som underminerar individers autonomi. Speciellt rekommendationssystem har identifierats som problematiska eftersom de kan utöva oönskad kontroll över människors liv. I denna uppsats försöker vi bedöma målen med design av rekommendationssystem för underhållningstjänster genom att göra en korpusanalys på 1 883 vetenskapliga artiklar om detta ämne. Vi genomför sedan en kvalitativ undersökning av psykologisk litteratur om koncepten *self-regulation, sense of agency* och *habits*. Dessa relaterar vi till användarautonomi. Vi kartlägger också relevant litteratur om användarcentrerad interaktionsdesign för att relatera uppfattningen om användarautonomi med användarvärde. Slutligen fokuserar vi på det specifika fallet för YouTubes rekommendationssystem och föreslår designändringar som syftar till att bättre anpassa tjänsteleverantörens mål till användarnas mål.

Vi drar slutsatsen att användarnas beteende är en felaktig återspegling av deras avsikter. På grund av detta riskerar användarens autonomi att undermineras av att endast beteendedata analyseras i rekommendationssystem. Många nuvarande rekommendationssystem, inklusive YouTube, använder beteendedata eftersom denna data är lätt att samla in och ofta lyckas maximera tjänsteleverantörens mål. Vi föreslår både korrigerande och förebyggande lösningar på detta problem. De korrigerande lösningarna fokuserar på att erbjuda användarna mer anpassningsbarhet. De förebyggande lösningarna fokuserar på sätt att samla in mer data som bättre motsvarar användarnas avsikter. Högre användaranpassning kan tillhandahålla användardata som kan förväntas motsvara användarens avsikt relativt bra.

Nyckelord

rekommendationssystem, autonomi, designetik, användarstudier, YouTube

Foreword

This thesis was done in cooperation by Anton Angwald and Kalle Areschoug.

Table of contents

1 Introduction

“We shape our buildings, and afterwards, our buildings shape us” - Winston Churchill

During the Trump presidency, references to George Orwell’s authoritarian dystopia *1984* became popular with both supporters and critics (Rodden, 2020). However, James Williams (2018), a tech-industry insider turned critic, proposes that people chose the wrong dystopian fiction writer. Williams argues that it’s more likely that we are approaching Aldous Huxley’s *Brave New World* (1932) incorporated by a Silicon Valley-fueled attention economy. In such an economy, where companies’ primary aim is to design products or services that compete for consumers’ attention (Goldhaber, 1997), Williams (2018) warns that we might face the Huxleyian problem of individual autonomy being undermined by entertaining distractions. This theme is taken to heart in the Netflix documentary *The Social Dilemma* which portrays recommendation algorithms as puppet masters, pulling the strings of a teenage boy with an iPhone-habit, by manipulating what content he is shown (Orlowski, 2020). While this portrayal might seem overdramatic, it relates to a problem for recommender system designers’ ability to successfully respect users’ autonomy when recommending content (Varshney, 2020).

Recommender systems have, in general, focused on maximising an engagement metric such as *“watching time”* and while if one user has watched one hour worth of music videos it seems reasonable to assume that this user values music, researchers have noted several issues with this approach (McNee et al., 2006; Pu et al., 2012; Knijnenburg et al., 2012; Nabizadeh et al., 2015; Ekstrand & Willemsen., 2016; Chen et al., 2019; Seaver et al., 2019). The central problem being, what in psychology is referred to as an *intention-behaviour gap* (Gollwitzer & Sheeran, 2006). Simply put, people do not always do what they intend to do. Behavioural user data such as *“watching time”* are also subject to *feedback loops* in recommender systems. Feedback loops are caused by the fact that the data being used to make recommendations is influenced by the recommendations themselves. This confounds the users’ intended behaviour with behaviour that might have been shaped by the system’s persuasive ability (O’Neil 2016; Covington et al., 2016; Mansoury et al., 2020). To come back to our previous example, the user that watched one hour of music videos might actually have wanted to study instead, but been too intrigued by the thumbnails of recommended music videos to do so. When finally quitting the application, perhaps the user did so with regret, feeling they wasted their time. This user would not be alone as previous studies have shown that it is common for users to complain that social media services waste their time and that they often later regret using it (Ames, 2013; Hiniker et al., 2016; Ko et al., 2015; Lukoff., 2018). We think the best way to approach public dissatisfaction with technological infrastructure, to assess its validity and propose solutions, is by applying it one-by-one, on a specific problem area and application. This is what we aim to do in this article. When it comes to the question of autonomy, we find recommender systems within the entertainment domain to be especially relevant to this type of analysis. Recommender systems technology can help us sort through a world of information abundance and help us make decisions, in ways that might support or undermine personal autonomy (Varshney, 2020).

Our main research question then, can be summarised as the following: “How can the design of recommendation systems for entertainment services align with the individual right to autonomy?” Our attempt at answering this question will be of exploratory nature. To deal with this question we will explore the following sub-questions:

1. How are problems of autonomy addressed in psychology, design research and philosophy?
2. Are the goals of recommendation systems aligned or in conflict with users’ autonomy?
3. How should research, in the fields of interaction design and recommender systems, address these problems?

To answer these questions we have conducted a broad literature overview on recommender systems research, followed by an in-depth qualitative review which will be the main focus of this thesis. During the course of our work we found that the issues we address are intertwined with various fields of research and rather than limiting our analysis to one field, we limit it to one use case, YouTube’s recommendation algorithm. This enables us to adopt an interdisciplinary approach which we hope might further bridge gaps to promote future interdisciplinary studies relating to user autonomy.

Since it might be difficult for the reader to intuitively understand the links between the various areas, we will start each section of the article by including a model portraying the topics that will be discussed. This model will gradually grow more complex and in the discussion section of this thesis we connect the various topics, utilising the full model. This introductory section includes the topics outlined in the figure on the next page.

Figure 1
Main topics of our introductory section

These topics will be further assessed, in the following part of the introductory section, in which we motivate our choice of YouTube as a specific use-case as well as give an introduction to previous research on the subject matter. We will conclude this introductory part of the thesis with the results of our broad literature overview.

1.1 Motivation

The practical importance of assessing whether recommender systems risk undermining individual autonomy by offering entertaining distractions is further stressed by public concerns with social media addiction (Hao, 2019; Nicas, 2018; Hornigold, 2019; Maack, 2019; Cullen, 2019). However, this criticism, along with other issues of user autonomy, are not solely linked to recommender systems. There are other influencing factors outside of recommender systems, in other types of interaction design (Gray, 2018) and especially in relation to the social psychological aspects of the network (Aksoy, 2018; Balakrishnan & Griffiths, 2017).

1.1.1 The case of YouTube

To reduce confounding social psychological factors, while still relating to a highly influential media platform, we will, when appropriate, focus on the specific case of YouTube's recommender algorithm. In comparison to Facebook, Instagram, Twitter or many other popular applications that also employ recommendation algorithms, YouTube is primarily an application focused on content rather than an application focused on social interaction. Even though the ideas outlined in this thesis can be extended to analysis of other recommendation systems than that of YouTube, focusing on YouTube has a value on its own since the platform has more than 2 billion monthly logged-in users, from more than 100 different countries. In total 1 billion hours of video are watched each day and more than 70% of this is on a mobile device (YouTube Press, 2021). At CES 2018 (a large tech industry event in Las Vegas), the Chief Product Officer at Google estimated in a panel discussion that YouTube's recommendation algorithm drives 70% of the watch time on the platform (Solsman, 2018). This suggests more specifically that the recommendation system has a persuasive effect on a large proportion of the global population. However, it would be misleading to discuss YouTube's recommendation algorithms in isolation from other design aspects on the platform and it can be argued that the recommender system as a whole is intertwined with the interaction design of the platform. Because of this we will also discuss other design aspects, even if the main focus of our discussion will be on recommendation algorithms.

1.1.2 Earlier research

The YouTube recommendation system has been shown to reduce users' sense of agency (Lukoff et al., 2021). This problem, relating to the broader problem of Autonomy & Personal identity have also been identified as one of six key areas of ethical concern in research on recommender systems in a literature overview by Milano et al (2020). Varshney (2020) has suggested that recommender systems might undermine sense of agency through relying too much on behavioural data in measuring the effectiveness of the system, he argues that it is essential to also include the notion of autonomy when evaluating recommender systems. Ekstrand and Willemsen (2016) also discuss the reliance on behavioural data in recommending content to users. The major advantage of implicit data is that it is more available, it consists of automatically collected data such as clicks and watching time. The other advantage is that it better predicts future behaviour in comparison to explicit ratings from users. According to Ekstrand and Willemsen the discrepancy between what users say (explicit data) and what users do (implicit data) can be explained either by *(a)* the user does not understand their true desires or *(b)* the user is dissatisfied with their behaviour and wishes to change it. The two options that Ekstrand and Willemsen propose provide an introduction to how the concept of user autonomy is intimately linked to user value. If *(b)* is true, optimising for users' behaviour instead of stated preferences can be reasonably argued to undermine autonomy, since it acts against the users own goals and wishes. However, if *(a)* is true, the objective for recommendation systems to retain individual autonomy becomes philosophically problematic. Should recommender systems aim to understand the “true” desires of users and optimise for these? Or is this a form of paternalistic stance that undermines personal autonomy by acting on the assumption that the individual is incapable of understanding their own best interests and therefore of taking their own decisions?

James Williams, dedicating a book to the topic of information technology and autonomy (2018), compares recommender systems to a GPS-system whose goal should be to guide us through digital space rather than through physical space. Entering an address on a GPS-system and ending up somewhere else would be evident of a faulty GPS-system and we should treat recommender systems by the same standard. Knijnenburg et al (2012) argued that the primary goal of recommender systems should be to increase user experience and that this is not the same thing as maximising prediction accuracy. They developed a framework for measuring user experience and conducted research that evaluates the connection between user experience and prediction accuracy. Their results give several examples of how there is a poor relationship between user experience and prediction accuracy. While prediction accuracy is measured implicitly, through for example clicks and interaction times, they measure user experience explicitly, through user testing of systems accompanied with interviews and surveys. Several other researchers have also challenged the assumption that algorithms which better predict behaviour lead to better recommender systems (Herlocker et al., 2004; Pu et al., 2012; Nabizadeh et al., 2015; Chen et al., 2019).

More than 2000 articles from 2011-2017 were surveyed by Singh et al (2021) in an overview on recommender systems and research directions. In this review they showed, among other things, that historically there has been a shift in recommender systems research from relying on explicit measurements to implicit measurements (Singh, Dutta Pramanik, Dey, & Choudhury, 2021). Based on their survey they also propose that future recommendation systems will with advent of the Internet of Things become truly ubiquitous and have positive impacts on our decision-making by giving us constant guidance in an information-saturated world. They say that *“The ideal RS [recommender system] should be like someone who knows us better than we know ourselves. They should sense our need and will suggest instinctively, even if we do not express it explicitly”.* A GPS-system that takes us to another location than the one we were planning to go to might undermine personal autonomy. Although, it is also what has been described by users as the good thing about YouTube’s recommendation system (Lukoff et al., 2021). Being able to discover content that users would not even know existed was sometimes valued as a meaningful exploratory journey.

The anthropologist Nick Seaver presents a gloomier outlook on recommendation systems in his article “Captivating Algorithms: Recommender Systems as Traps”.

He draws on fieldwork with recommender system designers to describe *"…a tendency among these systems' makers to describe their purpose as 'hooking' people – enticing them into frequent or enduring usage."* Reflecting on the change of explicit measurements to implicit measurements, Seaver says the following: *"Instead of predicting explicit ratings, developers began to anticipate implicit ones, and with this came a plainly captological approach to design: using traces of interactions recorded in activity logs, developers designed their systems to elicit more interactions. The prototypical recommender system was no longer a support for finding information, but a trap for capturing fickle users."* He suggests that there is a behaviourist paradigm within software design and traces its historical beginning to the Stanford Persuasive Technology Lab. Seaver argues that engagement metrics measure how successful a system is in capturing the users' attention but that there exists a conflict between developers and business people in which the former want to increase user satisfaction and the second want to increase user retention. However, since what is satisfying can also often be captivating, such a conflict can sometimes be resolved (Seaver, 2018).

1.2 Broad literature overview

The preceding articles helped us identify some key terms in relation to user value (preferences, satisfaction, implicit/explicit data) and in relation to user autonomy (prediction accuracy, engagement, capture, traps). We have also tried to establish the connection between these two main concepts.

To better understand the issues at stake and how current research practices might constrain possible solutions to these issues, we created our own text corpus on research articles on recommender systems ranging from those published 2008 to those published 2021. The main questions we wanted to address by analysing this corpus were:

1. Are claims that the majority of recommender systems (RecSys) research relies on implicit data substantiated? Or is it common for research to utilise both explicit and implicit data?
2. How common is user testing in the field?
3. Has the usage of terms like "user satisfaction", "user experience" or "preference" decreased after 2015? If this is the case, can this indicate that an increase in using behavioural data was also accompanied by a change in objective, aiming to predict user behavior rather than preference?

Since our objective is to discuss YouTube's recommendation system, what we learn from our research overview might not directly apply to this specific recommender system. However, there exists a substantive overlap between academic research and industry practice in recommender systems. This overlap is perhaps most visible when inspecting the proceedings of conferences and journals published by the Association of Computing Machinery (ACM), which is arguably the largest publisher of research material in this field. Many of these research outlets are closely followed, and have significant research contributions from the industry. Researchers from Google Inc (owner of YouTube) as well as researchers from Twitter have presented papers on their recommendation system at these conferences and we will look at these specific papers extensively in the next section. We hope that gaining a summarical view of the research field as a whole gives us a good starting point for later analysing these industry papers in more detail. We are also fond of the irony in addressing an issue of insufficient but easily quantified proxies within RecSys research by creating our own insufficient but easily quantified proxies for studying RecSys research.

1.2.1 Data collection

We collected and analyzed a total of 1,883 research articles to build a corpus of research concerned with media recommender systems. To collect the articles we used the following three databases; Springer, Scopus and ACM. These databases were chosen because of the high prevalence of research on recommender systems and since they were the ones used in the previously mentioned meta-study by Singh et al. (2021) (IEEE was excluded because of license restrictions on text-mining). While Singh only studied the abstracts of the collected articles, we gathered full articles to gain deeper insights. A weakness of our analysis is that due to issues with download restrictions, we have an uneven amount of articles from each year and publication. The final corpus consisted of 63% of articles from Springer, 20% from ACM and 17% from Scopus.

Table 1

Amount of articles per publication year

Publication Year	Number of Articles	Distribution
2021	63	3.3%
2020	317	16.8
2019	256	13.6
2018	233	12.4
2017	172	9.1
2016	142	7.5
2015	132	7
2014	104	5.5
2013	98	5.2
2012	74	3.9
2011	122	6.5
2010	62	3.3
2009	54	2.9
2008	54	2.9

We collected the data by searching for articles that within their abstract must contain the term "recommender systems" (or its synonyms) and that also contained at least one term related to media (video, music, news etc). We excluded abstracts that contained words like social media, health, medicine, tourism and e-learning since our analysis focuses on entertainment recommender systems.

A weakness of simply using all the articles that result from such a search is that it might both exclude relevant articles and include irrelevant articles because of general sentences in the abstract, such as *"Recommender systems have been used in a wide variety of areas such as in social media, music and medicine."* The exclusion of relevant articles is not a problem as long as it is not a systematic exclusion of which the systematicity is relevant to our thesis. Collecting articles manually would likely yield higher issues of systematic exclusion due to human biases. However, since the inclusion of irrelevant articles poses a larger issue, we choose search terms for which the results were rather relevant than plentiful. However, using such a large text corpora of articles in which the representative sample is close to the total number of topical research articles, anomalies and mistakes in the collected data become less important (Mayer-Schönberger, 2013).

1.2.2 Preprocessing of data

All the PDFs were uploaded to the software tool SketchEngine which automatically pre-processes the data. The PDF-files are converted to HTML-files and automatically annotated with grammatical metadata. We deleted the documents that contained a word count over 50 000. These documents were books, in which only a chapter was relevant to recommender systems. Duplicates were automatically deleted by SketchEngine. We used a combination of Linux command-line utilities to delete the reference-section of every article as well as further sort out irrelevant articles. For anyone interested in creating a searchable corpus of full scientific articles, we provide details on how we built ours in appendix 1.

1.2.3 Results

1. Are claims that the majority of RecSys research relies on implicit data substantiated? Or is it common for research to utilise both explicit and implicit data?

In order to address question *(1)* we needed to find proxy words for usage of implicit data and usage for explicit data, this effort proved to be unsuccessful. One promising candidate was to combine some variety of the articles consisting of participants with other articles using the word *"rating"* in specific contexts. However, the word rating was most regularly associated with the highly popular Movielens dataset, which consists of user ratings on movies but also includes some datasets on user-generated tags (metadata) on movies. The dataset is referred to in approximately 30% of the articles. Many of the articles that use the Movielens dataset or similar datasets are focused on implementing various mathematical models for improving the prediction of target data with training data. Generally, it is not important for these articles if the data is explicit rating data or implicit click data. The question of whether the researchers behind these articles utilise explicit or implicit data is therefore irrelevant. Another issue is the large number of discussion articles that widely discuss pros and cons of both types of data. We could try to approximate how these articles valued explicit data in comparison to implicit data by sentiment analysis, but this was outside of the scope of our work.

2. How common is user testing in the field?

Studying our corpus, we found that papers which included any form of user research that did not consist of automated data collection or online content-rating, had the text-string *"participant"* in them. The word is, however, also sometimes present in articles in which no study has been conducted. Out of the 2016 articles, 336 included the string *"participant"*. We performed a test in which we assessed a random selection of 30 of these articles and found that two out of these 30 articles did not include user research. Based on this, we approximate that 16% of research published on media recommendation systems include some form of user research.

3. Has the usage of terms like "user satisfaction", "user experience" or "preference" decreased since behavioural data started being used more than explicit data? Can this indicate that the rise in using behavioural data also included a conscious change in objective, aiming to predict user behaviour rather than preference?

Regarding question (3) we compared articles from 2008-2014 with articles from 2016-2021. We chose this time period since the period in between is when utilising implicit data became popular (Singh et al, 2021; Seaver, 2018). We saw that the number of occurrences of the term *"preference"* increased by 20% between these time periods, that *"satisfaction"* decreased by 23% and *"user experience"* decreased by 21%. In addition *"participant*" decreased by 21% and *"capture"* increased by 47%. However, it is hard to draw any conclusions from these observations, especially since we also saw that *"recommend"* decreased by 4% and *"system"* decreased by 9%. Since we expected these two terms to not have any noticeable differences between the periods, these results raise further questions about the validity of our measurement. Although except for *"preferences"*, the differences are quite large and follow a consistent trend that supports the argument that since the increasing reliance on behavioural data, more efforts have been focused on "capturing" users rather than increasing user utility.

Table 2

Percentage of change in the prevalence of certain terms. Change between the two periods 2008-2014 and 2016-2021.

Term	**Increase**	**Decrease**
Preference	20%	
Capture	47%	
Satisfaction		23%
User Experience		21%
Participant		21%
Recommend		4%
System		9%

Even if no conclusions could be drawn from our overview, it indicates that it could be worthwhile for future researchers to see if current RecSys research in comparison to previous RecSys research has consciously focused less on increasing user utility than on increasing engagement. This is the position that is taken in the social dilemma documentary (Orlowski, 2021) as well as Seaver's article (2018) and Williams's book (2018). If this view is correct, the main issue is an issue of incentive rather than a knowledge problem of using evaluation metrics in the right way.

Our broad literature overview provides clues to understanding the work practices and goals within recommender system research. It also shows the necessity to develop a qualitative understanding of how research within psychology, philosophy and design can help inform issues of user autonomy. This is what we will aim to do in the remainder of this article.

2 Main Section

We will now turn to the main part of our thesis, which consists of three subsections. In the following subsection we will focus on psychological concepts and studies related to user autonomy. In the second, we will do the same for ethics and design research and in the final we will assess YouTube’s recommendation algorithm.

2.1 Psychology of User Autonomy

Figure 2
Main topics of section: Psychology of User Autonomy

In this section we will look at the psychology of autonomy by three subcategories: self-regulation, sense of agency and habits. These three combined play a part of human autonomy. Later in the article we will look at how they are integrated with YouTube and technology.

2.1.1 Sense of agency

Autonomy is one of the main areas of ethical concerns regarding recommender systems but it is quite a broad concept. One psychological research article gives the following definition: *“Autonomy refers to self-government and responsible control for one's life.*” (Keller, 2016). One way of approaching autonomy without having to indulge in metaphysical debates on free will is to instead talk of a personal perception of autonomy; sense of agency. Sense of agency can be further divided into feelings of agency and judgement of agency (Moore, 2016). Feelings of agency is in-the-moment perception of agency and is linked to low-level sensorimotor processes. As an example, feeling in control when using an application, being able to click a button and seeing the interface react. Judgement of agency is a post-hoc perception of agency, estimating how much one was in personal control in a previous situation. It is linked to higher-level cognitive processes, integrating contextual information as well as background beliefs (Synofzik et al. 2013). Both of these refer to the subjective notion of being in control rather than actually being in control, sense of agency is therefore not the same as actual agency. They are, however, related to each other. For example, when interacting with technology, higher levels of automation leads to a decreased sense of agency (Moore, 2016). Klobas et al., (2019) conducted an interview study with participants that had self-reported a problematic bond with YouTube. One recurring denominator reported by almost all of the participants as a problem with YouTube, was in situations where the sense of agency decreased due to automation.

2.1.2 Self-regulation

One example where a user might have had high feelings of agency but low judgement of agency is when reflecting back, wondering why they just spent two hours on watching YouTube videos when they had the goal of spending just ten minutes before going to the store to buy groceries instead. This problem connects with self-regulation and problematic technology behaviour which can be described as a conflicting situation in which the user needs to inhibit desire in order to perform a goal-directed action. Such an inhibition of desire demands self-control. Repeatedly failing in resisting desire can result in learned helplessness, this refers to how individuals come to believe that they are unable to change the situation. This can lead to a negative spiral (Perry et al. 2010) since people in the state of learned helplessness, stop trying to resist desire. Hoffman et al. (2012) study how individuals with various levels of self control differ in how they manage to prevent themselves from acting on desires in conflict with their goals. Their study suggests that individuals with high levels of self control are not better at resisting desires, rather that they manage to avoid situations where conflicting desires are present in the first place.

In similarity to the studies from Hoffman et al. (2012), other studies support the thesis that individuals with lower inhibitory abilities as well as attentional control are at higher risk of suffering from what some researchers call *social-networks-use disorder*. Wegmann et al. (2020) defines "social-networks-use disorders*"* as a habitual usage triggered by impulsive responses to cues. These resemble addiction-like symptoms which include loss of control and repeated use, and continuing to use the application despite negative consequences. The authors write: *"A dominance of the impulsive system is assumed to induce approach tendencies towards potentially gratifying options while neglecting long-term risks, which may result in risky behavior such as drug consumption".* They suggest that the same approach tendencies can be a driving factor to using applications in a way that is later regretted, a problem that has been widely reported by users (Ames, 2013; Hiniker et al., 2016; Ko et al., 2015; Lukoff et al., 2018). Researchers on digital well-being have used the concept of *"lagging resistance"* to describe the self-reported tendency of users wanting to quit using an application but not wanting to do so just yet (Baumer et al., 2013, as reviewed by Lukoff et al., 2018). This concept relates to the conflict between instant gratification and long-term goal-achievement. There is, however, a debate in the research community whether social-networks-use-disorders and similar behavioural addictions should be classified as an addiction or if it should not (Mihordin, 2012; Billieux et al., 2014).

The concept of self-control mainly focuses on the ability to inhibit certain behaviour. This ability is part of the broader concept of self-regulation, which can be defined as the ability to regulate behaviour according to one's own goals (Hoffman et al., 2012). In similarity to the notion of learned helplessness, studies have shown that addicts experience a spiraling failure in self-regulation after what Marlatt called *"abstinence violation effect"* (Marlatt & Gordon, 2005 as reviewed by Heatherton & Wagner, 2012). This refers to how minor violations of self-regulative behaviour might lead to a total collapse of self-regulation. Other studies have shown that repeated situations in which self-regulation is required, depletes cognitive resources needed to exhibit self-regulation. Repeated exposure of tempting stimuli are thus likely to lead to failure in self-regulation. This concept is also referred to as ego-depletion (Marlatt & Gordon, (2005) as reviewed by Heatherton & Wagner, 2012; Wood et al., 2014). This is in accordance with the results from Hoffman et al. (2012) which suggest that individuals successful in self-control succeed by avoiding tempting situations in the first place rather than by having greater abilities to inhibit impulsive-behaviour. The same depletion of cognitive resources relevant for self-regulation has also been shown to be induced by information overload (Diamond, 2013). This is an aspect that recommender systems can and should help with, by helping the user sort through irrelevant information.

2.1.3 Habitual behaviour

Gollwitzer & Sheeran. (2006) found in their meta-analysis of meta-analyses that intentions only explained 28% of the variance in behaviour. They called this an intention-behaviour gap. Even though the exact size of the gap was difficult to estimate, the intention-behaviour gap was in the smallest estimates still large enough to show that it is common for people to behave against their own intentions. The authors suggested that the intention-behaviour gap might be due to habitual behaviour. In a dual processing framework, habitual behaviour can largely be described as fast, automatic and unconscious while non-habitual behaviour is slow, deliberate and reflective (Wood et al., 2014). In similarity to impulsive behaviour, habitual behaviour can function as a sequence of actions which are enacted as an automatic response to external stimuli (Wood et al., 2014). This is why Schnauber-Stockmann et al. (2018) say that habits function within the impulsive system, which can be contrasted to the reflective system. While processing within the reflective system has the disadvantage of relying on potentially effortful deliberation, it has the advantage of being oriented towards long-term goals and abstract values. The impulsive system is oriented towards immediate gratification and relies on cognitive heuristics which can be described as quick-and-dirty strategies for making decisions, highly susceptible to biases (Schnauber-Stockmann et al., 2018).

With this in mind, we can expect that negative media behaviour would largely be habitual behaviour. This is exactly what a user study on smartphone behaviour (Lukoff et al., 2018) suggested. They found that people sometimes experienced a loss of autonomy when they were using their smartphones and in these cases participants highlighted the habitual and automatic nature of their usage. Moreover, they found that *“Lack of control was rarely attributed to active failure to resist in-the-moment, but rather to unconscious habit.”* A similar finding comes from Van Deursen et al. (2015). They found that addictive smartphone behaviour was often associated with habitual smartphone use. The study by Lukoff et al. (2018) did not only show how habitual behaviour was a problem for user autonomy, it also showed that habitual behaviour was less valuable for users. This brings us to the next subsection in which we adopt an ethical design perspective to examine user autonomy in connection to user value. We use the term *“user value”* simply to denote the value that users gain from interacting with technology.

2.2 User Autonomy & User Value

Figure 3

Main topics of section: User Autonomy & User Value

In this section we will assess philosophical design literature in order to introduce the notion of user value and show how it is connected to user autonomy. This will inform a discussion on design proposals aimed at helping users with problematic technology behaviour.

2.2.1 What users really want

In the preceding section we have primarily discussed psychological concepts in relation to excessive social media usage. The results from the user study by Lukoff et al. (2018) not only show that habitual usage might lead to excessive usage, but also that it can be unsatisfactory in itself. Participants repeatedly reported habitual usage to be considered meaningless. As discussed in the previous section, dual process theories suggest that habitual actions can be in conflict with reflective processes. The prevalence of habitual usage might therefore reduce opportunities for reflection. These opportunities are crucial for turning attention to our own mental activities in order to *"call our beliefs and motives into question."* (Korsgaard, 1996 as reviewed by Williams 2018). Patterns of habitual usage might therefore also undermine the ability to form a personal identity which is connected to autonomy (Milano et al., 2020). Williams (2018) argues that in a competitive attention economy, it is easier for designers to appeal to our automatic, instinctual way of thinking rather than our reflective way of thinking. This is exemplified by a book on persuasive design by Nir Eyal's (2014) called *"Hooked: How to Build Habit Forming Products",* a book that is also mentioned in the article by Seaver (2018).

Williams (2018) concludes with the stance that *"Whether irresistible or not, if our technologies are not on our side, then they have no place in our lives."* but what "*on our side"* actually should mean is harder to define. Perhaps, we should not take for granted that reflective behaviour is necessarily better than instinctive behaviour. This taps into an active ethical debate on persuasive technology and nudging (Lyngs et al., 2018; Lembcke et al., 2019; Schmidt et al., 2020). General positions in this debate are characterised by Lyngs et al. (2018) in their *"...fictive dialogue between senior executives at a tech company aimed at helping people live the life they 'really' want to live".* Even if this fictive dialogue is quite an unorthodox way of approaching this ethical question, we find it to be of great pedagogical value for introducing the various positions in this debate for the unfamiliar reader. Below follows an excerpt from their paper, published at ACM's CHI conference:

Figure 4

Excerpt from '"So, tell me what users want, what they really, really want"' (Lyng et al., 2018)

But what does any of that actually mean? How can we be sure that we are giving users what they really want? What we need, my friends, is a clear answer to this question; a new metric towards which all our services should be geared; a new optimisation metric for life. So come on, hit me with your ideas!

Randy: I'm going to stop you right there, sir, if I may. What's wrong with our existing systems? We infer what users want from what they do and what other people like them do. If they spend every spare second watching cat videos, then our algorithms should give them more cat videos. If they keep watching them, that means our algorithms got it right. If they don't like them they will stop looking at them. Our algorithms will then show them less in the future ...

Harald: Woah there. I totally disagree. People are slaves to simple reward functions inherited from our evolutionary past. We know how to hack these reward systems, so if we leave people to their own devices (no pun intended) they will simply do whatever our algorithms nudge them to do. That might be binge-watching cat videos and ordering takeout pizza. It probably won't be filling in their tax returns or exercise ...

Nichola: But we could be nudging them to do those things instead! Even better, we could nudge them to do something truly worthwhile, like reading poetry, or contributing to science, or meditating on the miracle of their very existence!

By now it should be clear that correctly measuring user value is harder than predicting user behaviour. We could, like Nichola, base user value on assumptions of what is meaningful human behaviour. Although, this might backfire. The study by Lukoff et al. (2018) showed that even when users engage in productive or goal-directed behaviour, they experience this as meaningless if it is habitual. Persuasive interfaces aimed to nudge users to fulfill long-term goals might thus fail to increase user value if users are pushed to engage in these activities without reflection. Research by Hiniker et al. (2016) and Shin & Dey (2013) shows methods to detect when a user might be interacting habitually rather than intentionally. Inhibiting extensive app usage in those situations might better optimise user value. Another potential approach is proposed by Cheng et al. (2017). They construct a model that from two minutes of behavioural data can predict users' intention of the session. Ekstrand & Willemsen (2016) argue that explicit ratings (users' self-expressed desires) should continue to be included in the recommendation process alongside implicit data. In accordance with this, the study by Lukoff et al. (2018) shows that utilising explicit user ratings is a valid approach to measuring meaningfulness.

Perhaps the easiest way to avoid making assumptions about what users really want is by doing just this, asking the users themselves. At least in this case, the assumptions are the users' own assumptions. Therefore, even if it might not optimise user value it should at least optimise autonomy. A more recent study (Lukoff et al., 2021) did so qualitatively by asking users how YouTube's recommendation algorithm affects their sense of agency, as well as asking users what they thought of design proposals. 120 participants that were heavy users of YouTube participated in the study. Almost all of these participants found irrelevant recommendations to decrease their sense of control. For roughly half of the participants even relevant recommendations could decrease sense of control. Relevant recommendations decreased control in situations when the user was using the app habitually or at an unsuitable time (often late at night). The authors explained this issue as the result of recommendation algorithms being good at solving a *local optimisation problem* (*"what should I watch on YouTube?"*) while failing to solve a *global optimisation problem* (*"should I watch YouTube or not?"*). In relation to YouTube's recommendation system, Klobas et al. (2019) found that the behaviour of clicking on the linked related videos (that are chosen by recommendation algorithms) was strongly correlated with compulsive YouTube use. Even if sessions sometimes begin with users watching a goal-directed, productive video it becomes irrelevant to the original intention as the user clicks related recommended videos, each one deviating more from the starting video (Klobas et al., 2019). In the Lukoff et al. (2018) study on smartphone behaviour, users' intention was also gradually diminished which could be explained by design features promoting habitual usage.

2.2.2 Design proposals for helping users

In Lukoff's study of the YouTube platform, user participants expressed that currently available customisation settings helped in combating these problems but that they wanted more ability to customise the recommendations and the interface. However, the absolute majority of the participants were unaware of several of the already available customisation settings on YouTube. The authors' main suggestion, which stems from a small-scale participatory design study, is to include a customisable interface with various degrees of control. For example, enabling users to switch between a Focus mode and an Explore mode (Lukoff et al, 2021).

Another proposed solution is an intervention mechanism that will force the user to reflect over his/her usage. The user would be required to solve a cognitive task such as a puzzle in order to continue using the application. This might lead to combating the problem of habitual usage, as users are forced to become more aware of their usage (Park et al., 2018). Other researchers have proposed similar external mechanisms, such as enabling users to set the time of day or amount of time that they allow themselves to use an application (Hiniker et al., 2016).

There are other proposed solutions more directly concerned with changing recommender systems in order to avoid the need for a solution in the first place. One way in which this could be done is suggested in a paper from researchers affiliated with the Twitter platform at the 2021 ACM FAccT conference (Milli et al., 2021). The primary aim of their article is to directly respond to the issues outlined by Ekstrand & Willemsen (2016) by developing a more correct operationalisation of user value. Milli et al. (2021) combines different types of data by weighing them differently, in order to operationalise user value. Based on the assumption that explicit data better corresponds to user value they weigh data differently corresponding to how explicit it is. If the user clicks to view a tweet, this data has a low weight but if the user clicks the button "See less often" this data is given a high weight. However, regular performance metrics can not be used for approach as what is being optimised (user-value), can not be assessed directly, through the behavioural data that is fed into the model. Using measurement theory the authors argue that to evaluate user-value, first latent variables (or proxies) for user-value need to be validated. In accordance to the five forms of validity evidence as stated in the APA, the authors propose the following solutions on how to evaluate user-value:

1. Design platforms in a way that gives users the opportunity of giving rich feedback.
2. Conduct user research to empirically test whether user behaviour occurs for the reasons designers think they do. Do people click "See less often" because they want to see the tweet less often or accidentally because it is really close to the retweet button?

3. Observe whether internal structure functions as expected, see if theoretical assumptions hold
4. See interrelationships between variables. For example, by conducting explicit surveys to see if variables that are seen to be proxies for user value are correlated to proxies for user value taken from the explicit surveys.
5. Evaluate conceptualisations of user-value through consequences, do people stop using the platform or start to complain?

The authors successfully assess the validity of their judgements of how explicit a data point is through [*3*].

2.3 The YouTube Recommendation Algorithm

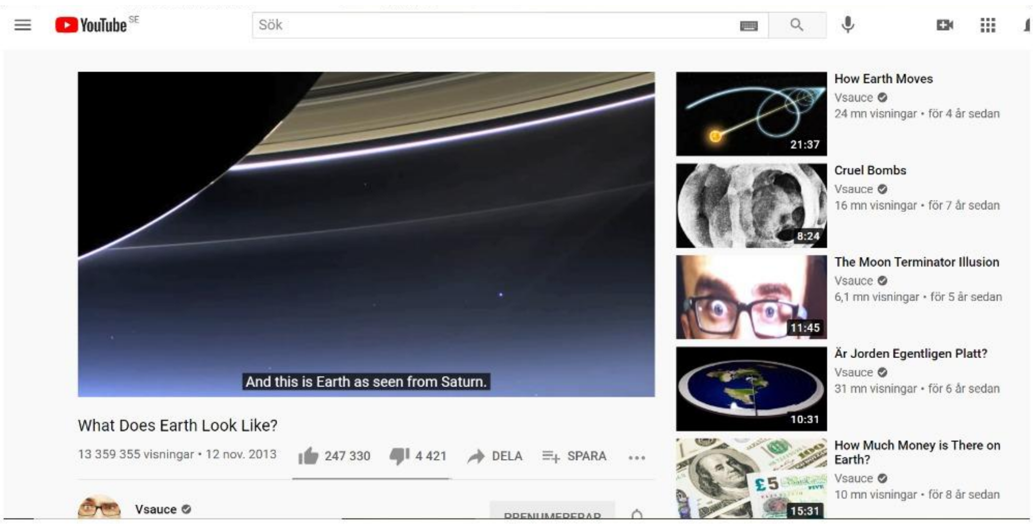
Figure 5

Main topics of section: The YouTube Recommendation Algorithm

We previously mentioned related videos that are linked to the video that a YouTube user is currently watching. These videos are chosen by the recommender systems by *(1)* combining how the recommended video relates to the currently watched video and by (2) choosing what video the user is likely to watch based on their history and similarity to other users (Zhao et al., 2019). Recommended videos are also presented at the home-page of YouTube, these videos are chosen on the basis of (2), while other sections on the home-page such as “Trending” are chosen on a different basis. The two articles that we discuss in this section mainly relate to linked related content but some aspects of it should also relate to the home page.

Figure 6

A video being played on Youtube, with related (recommended) videos in the right-hand section



At the 2016 ACM RecSys conference, researchers from Google presented a paper on how deep neural networks are utilised for YouTube recommendations (Covington et al., 2016). The authors describe that the recommendation of which video to watch next is generated through a candidate phase and a generation phase. In the candidate phase, hundreds of options are generated from the total video corpus consisting of many more, in the ranking phase these hundreds of videos are ranked. The objective of the system is to correctly predict the watch time for a video that might get recommended. The authors also note in their conclusion that watch time was increased dramatically in A/B testing, by utilising the methods outlined in the paper. The objective of the recommender system as a whole, therefore, seems to be to increase user watch time which the authors state is a promising proxy for engagement. The authors state in the beginning of the article that the only user data that is used, except for demographic data, is implicit data such as viewing time and click data. In the candidate phase, the types of data that are used can broadly be categorised as three types: semantic data, user profile data and behavioural collaborative-filtering data. The semantic data consists of data about the videos, such as keywords/tags, upload year and description. Examples of user profile data are demographic information and watch history. Behavioural collaborative filtering data contains information on which video other users have usually engaged with following the video the target user is currently watching (Covington et al., 2016).

Three years later, at the 2019 ACM RecSys conference, researchers from Google presented another paper regarding the YouTube recommendation system (Zhao et al., 2019). The Google researchers proposed and tested a Multi-gate Mixture-of-Experts (MMoE) system architecture. This system would take both engagement objectives and user satisfaction objectives into account. We propose, in line with previous discussion, that these are objectives in the interests of different parties, engagement objectives being of primary interest of the service provider and user satisfaction objectives of primary interest to the user. Now, there are several worthwhile points to note about YouTube's multi-objective recommendation system. We will now discuss these briefly before turning to them again in the next section, relating to the theoretical perspectives outlined in the previously discussed research.

1. First of all it is unclear how the two different objectives are balanced, and it seems like this, to a certain degree, is a choice up to the service provider. Why we say *"to a certain degree"*, is because the authors (Zhao et al., 2019) also indicate that it is limited by a lack of data related to user satisfaction (i.e video ratings and survey responses).
2. The multi-objective recommendation system is only utilised in the ranking phase. The candidate phase is the same as the one described in the 2016 article (Covington et al., 2016), which means that it still relies on implicit data (alongside demographic and semantic data).
3. The main reason that the MMoE is not used in the candidate phase is because of computational limitations, there are simply too many videos to evaluate in the candidate phase. The researchers note several times that the massive scale of the system limits the complexity of the models that can be used. Related limitations are that they are using a multimodal feature space (video tags, user demographics, thumbnails etcetera) and data availability.
4. Another major part of the article concerns combating position bias which can lead to feedback loops. Position bias concerns the issue that users are more likely to click recommended content that is highly ranked, engaging with this content increases the likelihood that it will be higher ranked for another user which creates a feedback loop. By being aware of this bias, researchers can combat this feedback loop by discounting the importance of

the engagement data on a video recommendation depending on the position of the video. In the 2016 article, the researchers do a similar thing with a bias towards older videos (Covington et al., 2016).

It is worth noting that we lack full information about this recommender system. Also, the system represented in these two articles might not represent the current YouTube recommendation system. It has been reported that YouTube makes changes to it's recommender system each year (Bergen, 2019). Therefore, it might be the case that changes that we propose to YouTube's recommender system are already in place, or non-applicable to the current system. The lack of transparency of how YouTube's recommender system functions has been criticised by researchers (Ranking Digital Rights, 2020; Singh, 2020) since it makes well-informed critiques of the system difficult. However, the point of basing our discussion on these two articles on Youtube's recommender system is to give technically feasible proposals for how user autonomy can be improved in general entertainment recommender systems.

3 Discussion

Figure 7
Main topics of discussion section

In this final section we will bring together the topics discussed in the main section with results from our broad literature overview as well as previous research on going beyond engagement metrics in evaluating recommender systems. We will first discuss how and why user- and service provider goals might become aligned and then give proposals for improving YouTube's recommender system. We will also offer a critical discussion of increasing user customisability, informed by the theoretical perspectives outlined in this article.

3.1 Aligning user goals and service provider goals

As previously stated, the optimisation goals for YouTube users can broadly be categorised in a local optimisation goal (finding the right video to watch) and a global optimisation goal (finding the right thing to do which might include either watching a video or doing something else). The service providers' optimisation goals can also be broadly categorised into

(1) an immediate goal of increasing engagement time in order to also increase ad revenue

(2) a long-term goal of avoiding reputational harms as well as recruiting and retaining a high number of users to ensure the platform stays influential

Problems for users can arise due to *(a)* the service providers' goals are in conflict with the user's goals or due to *(b)* the service provider fails to achieve goals. We can expect that the long-term goal *(2)* of the service provider is not necessarily in conflict with both of the users' goals, but the short-term goal *(1)* of the service provider is in conflict with the users' global optimisation goal. Even if the goals are not always in conflict with each other, the service provider can achieve their goals while not achieving the users' goals due to persuasive design taking advantage of irrational psychological tendencies. In the extreme case, the user might for example unwillingly watch videos that he/she dislikes for two hours because they are *"*captivating*"*.

Some of the research that we have surveyed suggests that the problem of aligning recommender systems to user goals is a problem of incentive (Orlowski, 2021; Seaver, 2018; Williams, 2018) and our broad literature overview also suggests this. At least part of the problem could then conceivably be that while the service providers can, they are reluctant to maximise user utility because it is in conflict with their own goals. The most promising solution would be something similar to legal restrictions that in one way or another force the service providers' goals to be aligned with the users' goals. However, we think that the long term goal (2) of service providers is fairly well aligned with user goals, and that there exists a strong will from service providers to resist resorting to legal restrictions. This can be part of the explanation why Mark Zuckerberg, the CEO of Facebook, announced on his Facebook page (2018) that they would from that year on *"make[ing] sure that time spent on Facebook is time well spent" (Zuckerberg, 12 jan 2018).*

The first thing we want to stress, even though it might seem obvious, is why big tech companies like Facebook and Google should pertain to users' long term goals. If the companies employ recommendation algorithms that have the ability to capture users for hours against their will, they risk these users deleting their accounts and quit using the platform. It is more worthwhile to keep an occasional user for 20 years than a heavy user for 20 hours. Retention should therefore be a more important metric than prediction accuracy of behaviour (Nabizadeh et al., 2015). It can be argued however, that the digital infrastructure that platforms such as Facebook offer is so reliant on *network effects,* that there are few competing platforms that the user can switch to. Network effect is an economic phenomenon that can refer to a service becoming valuable for a user depending on if enough people use the service. What's the point of being on a social network with only two people? It can be argued then that if a platform offers a large digital infrastructure that is essential to modern communication, users are unlikely to leave, even if they have bad experiences (Hughes, 1989).

YouTube, being a part of Google Inc, might want to resolve users' global optimisation problems for other reasons than to avoid reputational damage or to ensure that users are retained. Google has a wide array of services they provide. A user who's not watching videos on YouTube might use another of Google's services instead. The unprecedented size and influence of Google Inc also means that it has certain ethical responsibilities. We think both of these factors will become even more important if it is true that further development of the internet of things will lead to recommendation systems "*...becom[ing] truly ubiquitous and becom[ing] an essential tool in every sphere of* our life" (Singh et al., 2021).

Google, and similar companies have recently been under scrutiny during several US congressional hearings (D'onfro, 2018; Rushe & Paul, 2020; Feiner, 2021). Let us suppose that a similar future hearing would instead involve the issue of user autonomy and that because of the reasons stated in the preceding paragraph service providers say they are interested in optimising for users' values. They might, however, explain that they are not able to. Maximising user utility is easier said than done. First of all there exists the problem of finding the right technique, utilising the right type of data and the right type of evaluation metric. Maybe more pressingly, we need a definition of user utility and this leads us to having to define what is the "right thing to do" for any particular user, which traps us in the fictive dialogue of Lyng et al. (2018). Even if this question needs to be addressed eventually, systems can be improved for users while staying relatively agnostic about this question. This can be done by adopting the perspective that the users know best themselves what they want and that we therefore are best off trusting users' stated preferences.

3.2 User-centric design for Recommender Systems

The solutions proposed in Google Inc's own research papers (Covington et al., 2016; Zhao et al., 2019) and in the 2021 paper by Lukoff et al. as well as the experiments implemented on the Twitter platform show a promising step in the right direction. We will now critically assess these solutions by taking advantage of the theoretical background we have outlined in the earlier part of this paper.

Based on previous research (McNee et al., 2006; Gollwitzer & Sheeran, 2006; Pu et al., 2012; Knijnenburg et al., 2012; Ekstrand & Willemsen., 2016; Williams., 2018; Chen et al., 2019; Seaver et al., 2019; Milli et al., 2020; Varshney, 2020) we have shown that recommendations relying on implicit data can decrease user autonomy as compared to more explicit data. In light of this, it would be reasonable for YouTube to implement a similar approach to the one outlined in the article on Twitter's recommender system that we discussed in the previous section. Milli et al., (2020) utilises a weighting of different data points depending on how explicit they are. Data that is more explicit, such as user rating, is being valued higher than less explicit data, such as viewing time. Users' engagement data on a video that is retrieved by actively searching for it can be legitimately assumed to better represent users' intentions, as compared to data on a recommended video (Wood et al., 2014; Van Deursen et al., 2015; Park et al., 2018; Schnauber-Stockmann et al., 2018; Lukoff et al., 2018; Klobas et al., 2019; Lukoff et al., 2021). Because of this, the engagement data on a video retrieved from a user search query should be valued more than engagement data on a video retrieved from a recommendation. This might already be the case but due to a lack of transparency we are unsure.

YouTube should also extend their possibilities for explicit user feedback. This would decrease the problem of data sparsity for assessing user satisfaction. In the current video interface there are two major options for explicit feedback, a thumbs-up button and a thumbs-down button. This user feedback impacts recommendations (Covington et al., 2016; Zhao et al., 2019) but due to the ambiguous nature of these buttons there might exist a discrepancy between what the action actually means and what the designer expects it to mean (Milli et al., 2020). A simple solution to strengthening the validity of this explicit feedback is to replace these buttons with one for "Show more often" and one for "Show less often", the explicit instructions on what the button actually does can be expected to decrease the gap between what the action means and what the designer expects it to mean. This would also increase transparency by making it clearer for the user that the action impacts what videos are recommended. As supported by YouTube users' expressed desire for higher customisability (Lukoff et al., 2021), readily available buttons that clearly communicate an opportunity for user customisability might be used more often. Because of this, these buttons would not only provide more reliable data, they could also provide larger quantities of data, which would reduce problems of data sparsity.

YouTube (Zhao et al., 2019) proposes a multi-objective recommendation system which can balance the value of user satisfaction objectives and engagement objectives. The engagement objectives are of primary interest to the service provider and the satisfaction objectives are of primary interest to the user. The relative weight of each of these objectives can be modified. One could employ a recommendation algorithm that, as an arbitrary example, maximises user objectives 80% and service provider objectives 20% or that maximises each objective equally. The desired balance of these objectives is therefore a normative question that could perhaps be decided upon by an external auditor trained in ethical decision-making. However, if such an external auditor would decide that the relative weighting between objectives should be 50/50 (again, arbitrary example), we believe that the actual balance might still be skewed towards engagement objectives. The reason we believe this is due to a possible bias in the multi-objective recommendation system presented by Zhao et al., 2019, we will proceed to explain this bias in the following paragraph.

In the first phase (candidate generation phase), approximately 500 videos are generated from the total number of videos on YouTube (billions). These are then sorted in the second phase (ranking phase). Since the multi-objective system is difficult to implement on large quantities of data, it is not employed in the first phase. Instead, the first phase only utilises the engagement objective (alongside various semantic sorting) (Zhao et al., 2019). Because of this, the approximately 500 videos that are put into the second phase are videos with high engagement metrics, not necessarily videos that will satisfy the user. This means that in the second phase, the algorithm ranks a dataset that is already biased towards service provider (engagement) objectives over user satisfaction objectives. The most ethical solution according to us would be to revert this. Since both objectives can not be employed in the first phase, only the user satisfaction objectives should be used in the first phase. This might however prove difficult as the implicit data that can more readily be used to maximise engagement objectives is more plentiful. Utilising user satisfaction objectives, which rely on large amounts of explicit data, would lead to problems of data sparsity. This would likely result in poor recommendations both for user satisfaction and engagement. Instead the bias could be addressed in the same way as Zhao et al., (2019) addressed position bias. The bias towards engagement can be measured and discounted when balancing the two objectives in the second phase. In practice, this would mean always adding x amount to the desired weighting of the user satisfaction objective.

3.3 User Customisability

The two main problems that we have identified in relation user autonomy can be categorised as:

1. Excessive usage
2. Unsatisfactory usage

While unsatisfactory usage is more linked to local optimisation objectives, the two problems are entwined as a user might categorise something as excessive usage because it is unsatisfactory, they might also categorise something as unsatisfactory because it feels excessive. Since it is difficult to have an objective definition of what is satisfactory and what is not (Lyngs et al., 2018), these problems should be addressed from the perspective of user studies as well as discussions specifically set in the context of technology usage. The proposals outlined in the preceding section have mainly addressed unsatisfactory usage *(2).* If opportunities for explicit user feedback are only in-app, their satisfaction objective only covers local goals. Qualitative user studies like the ones we have previously surveyed (Lukoff et al., 2019; Lukoff et al., 2021) better address users' global optimisation problem.

When it comes to excessive usage, the psychological concepts we have previously discussed are highly relevant and designers should mitigate the risks for self-regulation failure in accordance with psychological literature. However, this might involve evaluating one type of user value over another. The concept of "lagging resistance" (Baumer et al., 2013) relates to the conflict between instant gratification and long-term goal achievement. A possible explanation to how this conflict might function that is consistent with the notion of learned helplessness (Perry et al., 2010) and the abstinence violation effect (Marlatt & Gordon, 2005 as reviewed by Heatherton & Wagner, 2012) is by a dissonance between feelings of agency and judgement of agency. Strong feeling of agency might be related to not being prevented from making an instantly gratifying choice and strong judgement of agency might be related to having successfully avoided instantly gratifying options in order to pursue long-term goal-directed behaviour. Therefore, designers might have to make the choice of optimising for sense of agency or judgement of agency. Some of the design proposals that Lukoff et al. (2021) suggest might reduce sense of agency while improving judgement of agency and this is also an issue of several possible design proposals such as lock-out mechanisms (Lukoff et al., 2021). This brings us into the muddy ethical territory of valuing one type of user agency over the other. Psychological literature suggests that a sense of agency can be tied to addictive behaviour but also that it has strong ties to satisfaction. When possible, solutions that do not directly decrease sense of agency therefore have the advantage of avoiding this trade off.

Another problem with this trade off, is that users are different. Several participants in the study by Lukoff et al. (2021) state that they use YouTube for different purposes at different times, sometimes wanting to be entertained and sometimes wanting to be able to focus. The psychological literature on social media addiction also suggests different needs for different types of people with regards to design. An addictive design feature for one person might simply be fun for another person and this also seems to hold for the same person but in different contexts. Because of this, Lukoff proposes high customisability for the user, one proposal being a *discover mode* and a *focus mode* in which the focus mode offers less or no recommendations. While we agree that high customisability is good we should think of the nature of that customisability. We propose that increased feelings of agency might lead to higher feelings of responsibility. If a user has a higher sense of agency and fails to exercise that agency, it leads to a higher self-attribution of that failure which in turn lowers the hindsight judgement of agency. This leads to the apparent contradiction that increased feelings of agency can in certain situations reduce the belief of self control abilities. Moreover, a reduction of this belief can lead to actual self-regulation failure (Perry et al., 2010; Marlatt & Gordon, 2005 as reviewed by Heatherton & Wagner, 2012).

External lock-out mechanisms are sensitive to this problem. An easily bypassed external lock-out mechanism that, for example, can be unlocked with a password, might both increase the users' feelings of agency and help the user exercise self-control. However, if the mechanism is easily bypassed, it is likely that the user will start to habitually ignore the lock-out mechanism and without reflecting, enter the password. The lock-out mechanism will then actually make it worse for the user because not only does it fail in helping the user to take a break, it also makes the user feel guilty for not taking a break. In the best scenario, this will lead to temporary ego-depletion. In the worst scenario, this behaviour over time might lead to learned helplessness and therefore reduce the users' self-regulation ability.

We think this problem, in giving the user higher customisability, should be addressed in two ways. First of all, it is essential that the user should not be given more choices if the choices are unlikely to actually make a difference to the user. If external lock-out mechanisms do not actually make significant reductions in compulsive behaviour they are more likely to do harm than good. Secondly, when customisability can make reductions in compulsive behaviour we propose that they can be framed in a way to increase user sense of agency while decreasing the likelihood of sense of guilt. We think that the fact that sense of agency does not perfectly correspond to actual agency can be utilised. One example of this is default bias. Having the default option being the current YouTube layout but having a continuously present option saying "enter focus mode", rather than having to choose between focus and entertain mode, gives different results for users. Only having the option to actively increase *"focus"* empowers users to make this choice while not introducing a sense of guilt for users who do not make this choice. An added benefit of such a user customisability is that YouTube gets one more point of explicit data. Knowing when people switch from the default mode to the focus mode is useful in understanding when people feel distracted by the recommendations. Analysing this data might give a better understanding of when users gain utility from recommendations and when they do not.

One external solution that is likely to bypass the problems mentioned in the two preceding sections is the one proposed by Park et al., (2018). They propose giving the user more autonomy by forcing them to reflect on their usage by inhibiting habitual behaviour through a cognitive task. We think this is likely to work but it is also likely to be annoying. This problem could perhaps be solved by optimising the difficulty and varying the type of task. However, while this option can be an empowering tool for users, it does not help aligning technology to users' goals in the first place. Aside from the fact that Williams (2018) makes a valid philosophical point that this should be a necessary ethical requirement for technology in the first place, it is also a temporary solution. If users need to employ empowering intervention mechanisms that deal with the problems of aversive technology, they need to be in a constant state of learning and discovering in order to be able to adapt to an ever-changing technological landscape.

4 Conclusion

It has been argued that individual autonomy risks being undermined by entertaining distractions created by an attention economy. We have highlighted how certain aspects of entertainment recommender systems can cause a problem for the individual autonomy of users. The primary problem we have discussed is how recommender systems try to predict the intentions of users from their behavioural data rather than from their expressed desires. Through assessing psychological literature on autonomy and user studies on entertainment services we have shown how users' behaviour is an inaccurate reflection of their intentions. We have also shown how only analysing behavioural data might undermine users' autonomy. Some of the research we have surveyed, as well as our broad overview of recommender system research, have indicated that the objective in using behavioural data has actually been to predict users' behaviour rather than their intentions. However, the evidence for this is far from conclusive and we have also discussed some reasons for why service providers might be interested in predicting users' intentions rather than only their behaviour. There are advantages to utilising users' behavioural data in recommender systems and problems with relying on users' expressed desires, mainly problems of data sparsity.

With this in mind, we have explored solutions to the following question:
"How can the design of recommendation systems for entertainment services align with the individual right to autonomy?"

These solutions have been both of preventive and corrective nature. The corrective solutions have been focused on offering users' more customisability. The preventive solutions have been focused on gathering more data that correspond better to users' intentions. We have also shown how higher customisability can provide user data that can be expected to correspond relatively well to users' intention.

Answering the question stated above will be a gradual undertaking and we have shown promising starting points for this venture. We have suggested that it is essential that users' right to autonomy is discussed in relation to users' values. This is to ensure that good-intentioned solutions aimed to increase users' autonomy do not result in unsatisfactory experiences.

If regulatory bodies are to demand more respect for user autonomy from recommender system designers, these demands must be specific. Judicial propositions should rely on a solid understanding of the constraints and possibilities of recommender system research. If the tech industry should be externally regulated or expected to regulate itself is up for debate. In any case, we think that an interdisciplinary understanding that connects the notions of user autonomy and value to recommender systems is the way forward in securing users' right to autonomy.

4.1 Future research

A specific suggestion for a future study is to examine the hypothesis outlined in our discussion section, that certain types of increased customisability can have negative effects on user autonomy. Such a study could be conducted by collecting and statistically comparing behavioural and explicit data from a control user group and a group using a manipulated YouTube interface that includes higher options for customisability. The study should assess whether:
A: changes in customisability leads to change in feelings of agency
B: changes in customisability leads to change in judgement of agency
C: changes in customisability leads to change in self-regulative abilities
D: There are interaction effects between *(A)*, *(B)* and *(C)*.

Utilising more conditions of customisability, researchers should assess *(A)*, *(B)* and *(D)* for customisability types that produce different effects on *(C)*. If our hypothesis is correct, customisability that seems to give the user more control but does not actually aid self-regulative abilities will produce a positive effect on (A), a negative effect on *(B)* and a negative effect on *(C)*. The data might have to be compared over a short (hours) and a longer time-period (weeks) to successfully describe the inter-relations between these variables.

Future research on digital autonomy should focus on building pedagogical bridges between interrelated subdisciplines within areas such as psychology, philosophy, interaction design, user studies and recommender system research. We have aimed to do this in our thesis but think that more work needs to be done. This work should provide the context required for in-depth user studies, in similar fashion to what was done by Lukoff et al. (2018) and Lukoff et al. (2021). We also think that future research should aim in innovating better methods for recommender systems to take users' values into account. This type of research should focus on a: finding useful, reliable proxies for users' values and b: develop methods for collecting explicit user ratings on larger scales.

5 References

Aksoy, M. E. (2018). A Qualitative Study On The Reasons For Social Media Addiction. European Journal of Educational Research, 7(4), 861-865. https://doi.org/10.12973/eu-jer.7.4.861

Ames, G. M. (2013). Managing Mobile Multitasking: The Culture of iPhones on Stanford Campus. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work (San Antonio, Texas, USA) (CSCW '13)*. ACM, New York, NY, USA, 1487–1498. https://doi.org/10.1145/2441776.2441945

Balakrishnan, J., & Griffiths M. D. (2017). Social media addiction: What is the role of content in YouTube? *J Behav Addict.* 2017 Sep 1;6(3):364-377. doi: 10.1556/2006.6.2017.058. Epub 2017 Sep 15. PMID: 28914072; PMCID: PMC5700732.

Baumer, E., Adams, P., Khovanskaya, V., Liao, T., Smith, M., Schwanda Sosik, V., & Williams, K. (2013). Limiting, leaving, and (re)lapsing: An exploration of Facebook non-use practices and experiences. *Conference on Human Factors in Computing Systems - Proceedings*. 3257-3266. 10.1145/2470654.2466446.

Bergen, M. (2019, August 1). YouTube Tweaked Algorithm to Appease FTC But Creators are Worried. *Bloomberg.* Gathered from: https://au.finance.yahoo.com/news/youtube-tweaked-algorithm-appease-ftc-151535936.html?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAEg15oRf38dfmvonLIieW08W4T414ME1VjJXC8cJHkd3_blGekAoPXC7qHQh9KmvwvVsEQBzdbSeCU698I5-8nYcI0gzRzu6LRTaB3iGGifOSuwoe1l8NrFVqwC3Qgaw8ooHxO1hfTG6UgCBmwoyKaxQ6SAMtORU6ceae9Wzjmmj

Billieux, J., Schimmenti, A., Khazaal, Y., Maurage, P., & Heeren, A. (2015). Are we overpathologizing everyday life? A tenable blueprint for behavioral addiction research. *J Behav Addict.* 2015 Sep;4(3):119-23. doi: 10.1556/2006.4.2015.009. Epub 2015 May 27. PMID: 26014667; PMCID: PMC4627665.

Billieux, J., Philippot, P., Schmid, C., Maurage, P., De Mol, J., & Van der Linden, M. (2014). Is Dysfunctional Use of the Mobile Phone a Behavioural Addiction? Confronting Symptom-Based Versus Process-Based Approaches. *Clin Psychol Psychother.* 2015 Sep-Oct;22(5):460-8. doi: 10.1002/cpp.1910. Epub 2014 Jun 19. PMID: 24947201.

Covington, P., Adams, J., & Sargin, E. (2016). Deep Neural Networks for YouTube Recommendations. In *Proceedings of the 10th ACM Conference on*

Recommender Systems (RecSys '16). Association for Computing Machinery, New York, NY, USA, 191–198. DOI:https://doi.org/10.1145/2959100.2959190

Chen, L., Yang, Y., Wang N., Yang, K., & Yuan, Q. (2019). How Serendipity Improves User Satisfaction with Recommendations? A Large-Scale User Evaluation. In *The World Wide Web Conference (WWW '19).* Association for Computing Machinery, New York, NY, USA, 240–250. DOI:https://doi.org/10.1145/3308558.3313469

Cheng, C., Lo, C., & Leskovec, J. (2017). Predicting Intent Using Activity Logs: How Goal Specificity and Temporal Range Affect User Behavior. In *Proceedings of the 26th International Conference on World Wide Web Companion (WWW '17 Companion)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 593–601. DOI:https://doi.org/10.1145/3041021.3054198

Cullen, D. (2019, May 3). YouTube addiction: binge watching videos became my 'drug of choice'. *The Guardian: Guardian Selects* (Medium Original). https://www.theguardian.com/us-news/2019/may/03/youtube-addiction-mental-health

Diamond A. (2013). Executive functions. *Annual review of psychology*, 64, 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

D'Onfro, J. (2018, december 11). Google's Sundar Pichai was grilled on privacy, data collection, and China during congressional hearing. *CNBC.* https://www.cnbc.com/2018/12/11/google-ceo-sundar-pichai-testifies-before-congress-on-bias-privacy.html?fbclid=IwAR0NhgcBL2FaNQQCB_eP-r_WGKM__nlUVpY82I7O-mL4p1CSS8wAEyFUl_A

Goldhaber, M. H. (1997). The attention economy and the Net. *First Monday*, 2(4). https://doi.org/10.5210/fm.v2i4.519

Gollwitzer, P., & Sheeran, P. (2006). Implementation Intentions and Goal Achievement: A Meta-Analysis of Effects and Processes. First publ. in: *Advances in Experimental Social Psychology 38* (2006), pp. 69-119. 38. 10.1016/S0065-2601(06)38002-1

Gray, M. C., Kou, Y., Battles, B., Hoggatt, J., & Toombs, L.A. (2018). The Dark (Patterns) Side of UX Design. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, Paper 534, 1–14. DOI:https://doi.org/10.1145/3173574.3174108

Ekstrand D. M., & Willemsen C. M. (2016). Behaviorism is Not Enough: Better Recommendations through Listening to Users. In *Proceedings of the 10th ACM Conference on Recommender Systems (RecSys '16)*. Association for Computing Machinery, New York, NY, USA, 221–224. DOI:https://doi.org/10.1145/2959100.295917

Eyal, N., & Hoover, R. (2014). *Hooked: How to build habit-forming products.*

Feiner, L. (2021, march 24). Here's what the CEOs of Facebook, Google and Twitter will tell Congress about fighting misinformation. *CNBC.* https://www.cnbc.com/2021/03/24/facebook-google-and-twitter-ceos-return-to-congress-heres-what-theyll-say.html?fbclid=IwAR2Dd53SsIUe8FAzSpHl449vqjvm6wuflB0FaLNOXlU3PHzg9oiHq06R-Bs

Flayelle, M., Maurage, P., & Billieux, J. (2017). Toward a qualitative understanding of binge-watching behaviors: A focus group approach. *J Behav Addict.* 2017 Dec 1;6(4):457-471. doi: 10.1556/2006.6.2017.060. Epub 2017 Oct 13. PMID: 29025269; PMCID: PMC6034945

Hao, K. (2019, September 27). "YouTube is experimenting with ways to make its algorithm even more addictive". *MIT Technology Review.* https://www.technologyreview.com/2019/09/27/132829/youtube-algorithm-gets-more-addictive/

Heatherton, T. F., & Wagner, D. D. (2011). Cognitive neuroscience of self-regulation failure. *Trends in cognitive sciences*, 15(3), 132–139. https://doi.org/10.1016/j.tics.2010.12.005

Herlocker L. J., Konstan, A. J., Terveen, G. L., & Riedl, T. J. (2004). Evaluating collaborative filtering recommender systems. *ACM Trans. Inf. Syst.* 22, 1 (January 2004), 5–53. DOI:https://doi.org/10.1145/963770.963772

Hiniker, A., Hong, R. S., Kohno, T., & Kientz, A. J. (2016). MyTime: Designing and Evaluating an Intervention for Smartphone Non-Use. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 4746–4757. https://doi.org/10.1145/2858036.2858403

Hofmann, W., Baumeister, R., Förster, G., & Vohs, K. (2012). Everyday temptations: an experience sampling study of desire, conflict, and self-control. *Journal of personality and social psychology*, 102 6, 1318-35.

Hornigold, T. (2019, October 17). "Algorithms Are Designed to Addict Us, and the Consequences Go Beyond Wasted Time". *SingularityHub*. https://singularityhub.com/2019/10/17/youtubes-algorithm-wants-to-keep-you-watching-and-thats-a-problem/

Hughes, T. (1989) The Evolution of Large Technological Systems. In: Bijker, W.E., Hughes, T.P. E Pinch, T.J. The Social Construction of Technological Systems. *New Directions in the Sociology and History of Technology.* (pp. 51-82) Massachusetts: MIT Press

Huxley, Aldous. Brave New World. New York: *Harper Brothers,* 1932. Print.

Keller, H. (2016). Psychological autonomy and hierarchical relatedness as organizers of developmental pathways. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 371. 20150070. 10.1098/rstb.2015.0070.

Klobas, J., McGill, T., Moghavvemi, S., & Paramanathan, T. (2018). Problematic and extensive YouTube use: first hand reports. *Online Information Review*. 43. 10.1108/OIR-01-2018-0032.

Knijnenburg, B. P., Willemsen, M. C., Gantner, Z., Gantner, Z., Soncu, H., & Newell, C. (2012). Explaining the user experience of recommender systems. *User Model User-Adap Inter* 22, 441–504 (2012). https://doi.org/10.1007/s11257-011-9118-4

Ko, M., Yang, S., Lee J., Heizmann, C., Jeong, J., Lee U., Shin D., Yatani, K., Song, J., Chung, K.-M. (2015). NUGU: A Group-based Intervention App for Improving Self-Regulation of Limiting Smartphone Use. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Vancouver, BC, Canada) (CSCW '15). ACM, New York, NY, USA, 1235–1245. https://doi.org/10. 1145/2675133.2675244

Korsgaard, C. M., & O'Neill, O. (1996). The sources of normativity. Cambridge: *Cambridge University Press*.

Lembcke, T.-B., Engelbrecht, N., Brendel, A. B., & Kolbe, L. (2019). "TO NUDGE OR NOT TO NUDGE: ETHICAL CONSIDERATIONS OF DIGITAL NUDGING BASED ON ITS BEHAVIORAL ECONOMICS ROOTS" *Ethical Considerations of Digital Nudging/Twenty-Seventh European Conference on Information Systems (ECIS2019),* Stockholm-Uppsala, Sweden.

Lukoff, K., Yu, C., Kientz, J., & Hiniker, A. (2018). What Makes Smartphone Use Meaningful or Meaningless? *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol*. 2, 1, Article 22 (March 2018), 26 pages. DOI:https://doi.org/10.1145/3191754

Lukoff, K., Lyngs, U., Zade, H., Liao, J., Choi, J., Fan, K., Munson, S., Hiniker, A. (2021). How the Design of YouTube Influences User Sense of Agency. ArXiv, abs/2101.11778.

Lyngs, U., Binns, R., Van Kleek, M., & Shadbolt, N. (2018). "So, Tell Me What Users Want, What They Really, Really Want!". In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems (CHI EA '18)*. Association for Computing Machinery, New York, NY, USA, Paper alt04, 1–10. DOI:https://doi.org/10.1145/3170427.3188397

Maack, M. (2019, June 14). 'YouTube recommendations are toxic,' says dev who worked on the algorithm. *The Next Web*. https://thenextweb.com/news/youtube-recommendations-toxic-algorithm-google-ai

Mansoury, M., Abdollahpouri, H., Pechenizkiy, M., Mobasher, B., & Burke, Robin. (2020). Feedback Loop and Bias Amplification in Recommender Systems. In Proceedings of the 29th *ACM International Conference on Information & Knowledge Management (CIKM '20)*. Association for Computing

Machinery, New York, NY, USA, 2145–2148. DOI: https://doi.org/10.1145/3340531.3412152

Marlatt, G. A., & Gordon, J. R. (1985). *Relapse prevention: Maintenance strategies in the treatment of addictive behaviors.* New York: Guilford Press

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work and think.* London: John Murray.

Mihordin, R. (2012). Behavioral Addiction—Quo Vadis?. *The Journal of Nervous and Mental Disease*, 200(6), 489–491. doi: 10.1097/NMD.0b013e318257c503

McNee M.S., Riedl, J., & Konstan, A.J. (2006). Being accurate is not enough: how accuracy metrics have hurt recommender systems. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems (CHI EA '06). Association for Computing Machinery*, New York, NY, USA, 1097–1101. DOI:https://doi.org/10.1145/1125451.1125659

Milli, S., Luca, B., & Moritz, H. (2020). "From Optimizing Engagement to Measuring Value". arXiv:2008.12623

Moore J. W. (2016). What Is the Sense of Agency and Why Does it Matter?. *Frontiers in psychology*, 7, 1272. https://doi.org/10.3389/fpsyg.2016.01272

Nabizadeh, A. H., Jorge, A., & Leal, J. (2015). Long Term Goal Oriented Recommender Systems. 552-557. 10.5220/0005493505520557.

Nicas, J. (2018, February 7). "How YouTube Drives People to the Internet's Darkest Corners". *The Wall Street Journal.* https://www.wsj.com/articles/how-youtube-drives-viewers-to-the-internets-darkest-corners-1518020478

Milano, S., Mariarosaria, T., & Luciano, F. (2020). Ethical aspects of multi-stakeholder recommendation systems. *The Information Society*. 37. 1-11. 10.1080/01972243.2020.1832636.

O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown Publishing Group, USA.

Orlowski, Jeff. (Director). *The Social Dilemma*. [Film]. Netflix, Netflix.com, 9 Sept. 2020, https://www.netflix.com/title/81254224

Ranking Digital Rights. (2020). *RDR Corporate Accountability Index: Transparency and accountability standards for targeted advertising and algorithmic systems PILOT STUDY AND LESSONS LEARNED.* Open Technology Institute. https://rankingdigitalrights.org/wp-content/uploads/2020/03/pilot-report-2020.pdf

Park, J. Y., Sim, J. Y., Kim, J., Yi, M. Y., & Lee, U. (2018). Interaction Restraint: Enforcing Adaptive Cognitive Tasks to Restrain Problematic User Interaction. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems (CHI EA '18).* Association for Computing Machinery, New York, NY, USA, Paper LBW559, 1–6. DOI:https://doi.org/10.1145/3170427.3188613

Perry, R., Chipperfield, J., & Stewart, T. (2010). Perceived control. https://www.researchgate.net/publication/316747941_Perceived_control

Pu, P., Chen, L., & Hu, R. Evaluating recommender systems from the user's perspective: survey of the state of the art. *User Model User-Adap Inter* 22, 317–355 (2012). https://doi-org.ezproxy.ub.gu.se/10.1007/s11257-011-9115-7

Rodden, J. The Orwellian "Amerika" of Donald J. Trump?. *Soc 57*, 260–264 (2020). https://doi.org/10.1007/s12115-020-00476-5

Rushe, D., & Paul, K. (2020, july 19). 'Too much power': Congress grills top tech CEOs in combative antitrust hearing. *The Guardian.* https://www.theguardian.com/technology/2020/jul/29/tech-hearings-facebook-mark-zuckerberg-amazon-jeff-bezos-apple-tim-cook-google-sundar-pichai-congress?fbclid=IwAR3dUkcmHsh_ytglw_vO1ecoXOgh9JMGKsBWlQWrdgrhg95YDbqDSncZTLI

Schmidt T. A., & Engelen, Bart. (2020). "The ethics of nudging: An overview" *Philosophy compass*. Vol. 15: 4. https://doi.org/10.1111/phc3.12658

Schnauber-Stockmann, A., Meier, A., & Reinecke, L. (2018). Procrastination out of Habit? The Role of Impulsive Versus Reflective Media Selection in Procrastinatory Media Use. *Media Psychology.* 21. 1-29. 10.1080/15213269.2018.1476156.

Seaver, N. (2019). Captivating algorithms: Recommender systems as traps. *Journal of Material Culture.* 2019;24(4):421-436. doi:10.1177/1359183518820366

Shin, C., & Dey, A. (2013). Automatically detecting problematic use of smartphones. UbiComp 2013 - *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing.* 10.1145/2493432.2493443.

Singh, S. (2020). Why am I seeing this? - Case Study: YouTube. *Open Technology Institute.* https://www.newamerica.org/oti/reports/why-am-i-seeing-this/case-study-youtube

Singh, P., Pijush, D. P., Avick, D., & Prasenjit, C. (2021). Recommender Systems: An Overview, Research Trends, and Future Directions. *International Journal of Business and Systems Research.* 15. 14–52.

Solsman, J. (2018, January 10). Youtube's AI is the puppet master over most of what you watch. *CNET.* https://www.cnet.com/news/youtube-ces-2018-neal-mohan/

Synofzik, M., Vosgerau, G., & Voss, M. (2013). The experience of agency: an interplay between prediction and postdiction. *Frontiers in psychology*, 4, 127. https://doi.org/10.3389/fpsyg.2013.00127

Van Deursen J.A.M. A., Bolle, L. C., Hegner, M. S., & Kommers, A.M. P. (2015). Modeling habitual and addictive smartphone behavior: The role of smartphone usage types, emotional intelligence, social stress, self-regulation, age, and gender, *Computers in Human Behavior, Volume 45*, 2015,Pages 411-420, ISSN 0747-5632, https://doi.org/10.1016/j.chb.2014.12.039.

Varshney, L.R. (2020). Respect for Human Autonomy in Recommender Systems. ArXiv, abs/2009.02603.

Wegmann, E., Müller, S.M., & Turel, O. (2020). Interactions of impulsivity, general executive functions, and specific inhibitory control explain symptoms of social-networks-use disorder: An experimental study. *Sci Rep 10*, 3866 (2020). https://doi.org/10.1038/s41598-020-60819-4

Williams, J. (2018). *Stand out of our Light: Freedom and Resistance in the Attention Economy*. Cambridge: Cambridge University Press. doi:10.1017/9781108453004

Wood, W., Labrecque, J. S., Lin, P.-Y., & Rünger, D. (2014). Habits in dual-process models. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual-process theories of the social mind* (p. 371–385). The Guilford Press

Youtube official blog. (2021). Youtube by the numbers. Retrieved 2021-05-12 from https://blog.youtube/press/

Zhao, Z., Hong, L., Wei, L., Chen, J., Nath, A., Andrews, S., Kumthekar, A., Sathiamoorthy, M., Yi, X., Chi, E. (2019). Recommending what video to watch next: a multitask ranking system. In *Proceedings of the 13th ACM Conference on Recommender Systems (RecSys '19)*. Association for Computing Machinery, New York, NY, USA, 43–51. DOI:https://doi.org/10.1145/3298689.3346997

Zuckerberg Mark. (2018, january 12). make[ing] sure that time spent on Facebook is time well spent. [Facebook update]. https://www.facebook.com/zuck/posts/one-of-our-big-focus-areas-for-2018-is-making-sure-the-time-we-all-spend-on-face/10104413015393571/

6 Appendices

Appendix 1:

How to create a corpus of scientific articles in SketchEngine

Search and download articles
Search for articles with a specific search query in a database. Use the Chrome-plugin *“Simple Mass Downloader”* and sort by the metadata name of the website’s PDF-links in order to download all linked PDF-files. Placing them in separate folders for year and publication

Convert to text with SketchEngine
Zip the folders and upload to SketchEngine. Unpack the folders in the file-handler of SketchEngine and add attribute year as well as publication to the files in each folder. Download Corpus as a text document.

Linux command-line
Open the Linux command-line and change directory to where the downloaded text file is stored and enter the commands below.

Split downloaded corpus-file to independent text-files:
csplit remainspinr21.txt '/</doc>/' '{*}'

Delete reference section of all papers
sed -i '/^References\|^references/,$d' *

Copy files to back-up folder, containing keyword x times
grep -o -c -i 'keyword\|or_diff_keyword' * | awk -F: '{if ($2 > x){print $1}}' | xargs cp -t /home/x_location/

Delete files with keyword occurring x times or more:
grep -o -c 'keyword\|or_diff_keyword' * | awk -F: '{if ($2 > x){print $1}}'| xargs rm

Add .txt extension to all files in folder:
rename 's/$/\.txt/' *

Zip the folder containing the files and upload to SketchEngine