Should Personalization Be Optional on Paid Streaming Platforms? An Experiment on User Preferences for Personalization or Increased Data Privacy

Bengusu Ozcan

Abstract: Users' willingness to pay for personal data privacy or sell their personal data to big internet platforms leveraging online advertising have been researched extensively. Despite not being researched as much, paid streaming platforms, such as Netflix or Spotify, also utilize personal data to enable their personalized services. Unlike receiving personalized ads from free internet services, users of these platforms pay for the service and also enable the company to use their personal data to receive personalized recommendations. However, there are not many studies comparing the value the user receives from personalization versus the value the platform extracts from users' data. We conducted an online experiment assuming that if people knew the extent to which their data is utilized by paid streaming platforms, their willingness to pay for these services may change, hence these platforms may consider personalization as an optional feature. Our results did not find strong evidence supporting this hypothesis. Our findings indicate that users may not reflect their true behavior in studies about online personal data preferences, which is addressed by acceptability gap or privacy paradox in the literature.

Key Terms: personalization algorithms, data privacy, personal data, acceptability gap, privacy paradox, online streaming, online advertising, data economy, data ownership, digital services

A thesis presented for the degree of
Master of Arts

Quantitative Methods in the Social Sciences
Columbia University December 2021

Contents

List of Figures

List of Tables

1 Literature Review

1.1 Online Advertising and Personal Data Exchange

Big tech companies like Facebook, Google and Amazon, have been in the center of public and legislative debates from online data privacy discussions to antitrust. As exemplified by the Cambridge Analytica Scandal, in which Facebook allowed a third party to access their users' friends' data for political campaigning and potentially impacted 2016 US Elections (Confessore 2018), the power of targeting certain groups and individuals online has started to raise more concerns. The main concern is that these platforms allow third parties to use this targeting power for online advertising and sponsored content, which is observed to be used for malicious purposes as well such as political propaganda or racial discrimination (Romm 2020).

To give a brief snapshot on how online advertising works, platforms like Facebook, Google or Amazon collect online information about their users and segment them based on the attributes and tendencies that this data reveals about them. As shown in Vigderman and Turner (2021)'s snapshot of what data big tech companies collect, depending on whether the user is registered or not, this can be personally identifiable information (PII) that reveal a certain individual, such as their name or contact details, as well as demographic information that does not reveal the identity but the profile of the user such as gender or age and finally their online activity information (OAI) such as keywords they searched or videos they watched. In this paper, I will refer to all the data a user generates on online platforms as their online personal data (OPD), which includes PII, demographic data and OAI. This data is leveraged by third parties to place advertised content on these platforms by targeting specific audiences, such as women with kids in a certain country, or people who searched for certain political keywords. As a common business model, companies using OPD often offer their services for free, such as search engines or social media platforms, in exchange for monetizing OPD by online advertising or sponsored content. The algorithms processing OPD for these purposes are not visible to users and work behind the scenes while the service is used. (Gran, Bucher, and Booth 2020). In other words, the user generates value for these platforms through this data exchange without being fully aware of it since the process is not as explicit as a commercial transaction in which the user sees the amount of money they need to pay and the equivalent service they would get and confirm the transaction.

1.2 Getting Paid for Sharing Data vs Paying for Data Privacy

The implicit value exchange between these companies and the users seems rather simple: allow the company to monetize your data and access the services for free. However, there are indications that users' OPD may be more important to them than the benefits they receive from an online service, which we will introduce in this section. Two approaches to address this concept are compensating users for sharing their data or on the contrary, users to pay online services for using those services without sharing their data.

First approach about this value exchange brought forth by scholars such as Tim Wu (2015) is whether companies would owe users a share of their profits earned from user data in cases where companies experience disproportionately greater overall benefit compared to the benefit users receive from accessing these services for free. This suggestion may sound like a fair exchange that would incentivize companies to minimize the amount of data they collect only to absolutely necessary ones, but it brings operational difficulties such as estimating the value that the user data generates for the company or its subjective sensitivity for the user, especially for different data types ranging from PII to OAI (Zylberberg and Ott 2016). A more direct rejection to this argument comes from Electrical Frontiers Foundation (EFF), suggesting that estimating a monetary value for user data undermines individuals' right to their full data privacy, which EFF defends should be recognized as a fundamental human right (Tsukayama 2020).

Another approach is quite the opposite of being compensated for sharing data, but rather paying for the data privacy itself, which has led to many studies in recent years. Prince and Wallsten found in a survey experiment across Germany, USA and 4 Latin American countries that users expressed willingness to pay for privacy of their PII and demographic information on various online platforms, and their willingness to pay did not change when respondents were reminded explicitly about the benefits they receive from the platforms in exchange of personal data (2020). Another study on a UK sample about how willingness to pay for OPD privacy changes depending on the OPD type found a strong reverse correlation between importance of data and willingness to share (Maple et al. 2019). Finally, a study conducted on a US sample asked about users' willingness to pay for privacy of their OPD and also how much they would demand a company to pay them to access their OPD, across different OPD types including PII and demographic data such as age, sex, income and political affiliation. Interestingly, for both questions, the survey found that people valued their demographic data more than their PII, although demographic data does not reveal their identity as a sensitive information that PII contains (Angela and Sunstein 2019).

All these studies about being paid for OPD or willingness to pay for the privacy of OPD indicate that all types of OPD, regardless of being PII, OAI or demographic data, has a subjective level of sensitivity and importance to the user. This entails a need for users to be fully informed about the extent to which their OPD would be used in order to enable them to evaluate the benefit they get from sharing their OPD.

1.3 Acceptability Gap Between Personalization and Data Sharing

A common criticism to both the ideas of platforms paying users for accessing their OPD or users paying platforms for the privacy of their OPD comes from a concept that is referred to as the "acceptability gap". This concept explains that users miss the precedence relationship between providing their OPD and receiving personalized services which is essential for users to have a practical benchmark comparing the value received from personalization to the value of OPD. A survey conducted across the UK, Germany and the US found that the majority of the people opposed the collection of their OPD by digital services, however the majority also found personalized services acceptable (Herzog et al. 2021). The study further proposed that the acceptability gap may occur since users consider providing their OPD as an inevitable trade-off. In other words, users may simply think that their OPD would be collected no matter what, hence, they had better receive personalization as a benefit at least. A connected concept to acceptability gap is the so-called privacy paradox, which is explained by Barth and De Jong as (2017):

> "... discrepancies [exist] between user attitude and their actual behavior. More specifically: While users claim to be very concerned about their privacy, they nevertheless undertake very little to protect their personal data."

The privacy paradox has also been criticized as a concept, because the reported online behavior can diverge a lot from the actual. For example, it is tricky to assume that people's attitudes towards OPD privacy are applicable to their decisions in daily life because their report in a survey setting may be amplified or anchored. (Kokolakis 2017).

To sum up, there is no consensus on whether users should receive compensation for sharing their OPD or pay for increasing the privacy of their OPD in the literature. Given the concepts such as acceptability gap that addresses a lack of user understanding in how OPD is processed by online platforms, users may need more transparent and explicit information to evaluate the exchange between their OPD and the equivalent service, and eventually have the option of opting in for or out of this exchange.

1.4 Paid Streaming Platforms and Their Benefit from Online Personal Data

We have discussed the utilization of OPD and users' preferences on OPD privacy primarily for digital platforms that provide their services for free. However, this paper will focus specifically on paid online streaming platforms (PSPs), which provide their services by charging the users with a subscription fee, such as Netflix or Spotify Premium (hereinafter Spotify). Given that these platforms do not primarily rely on the online advertising model, studies about OPD privacy on PSPs are not as abundant as other digital services, however these platforms also have a structure of benefiting from their users' OPD without the users being explicitly informed. Moreover, despite charging the users for their service, these platforms collect and process OPD by default without providing their users an option to opt out. This makes PSPs particularly insightful to explore the value of OPD, because there is already a direct monetary exchange between the user and the platform, which may provide a more measurable basis for the user to assess the value of their OPD. From now on, we will focus on PSPs by explaining how they utilize OPD and user awareness in the process.

While most PSPs provide high-level explanations about how their personalization algorithms work, they do not explicitly reveal the inner workings of these algorithms in the interest of intellectual property protection. For example, Netflix's official customer support website has a section which outlines how the personalized recommendations work, and lists the following as a factor taken into account: "other members with similar tastes and preferences on our service." (Netflix, n.d.) Spotify does not provide a similar explanation on their official website, however a few external technical resources claim that Spotify uses a similar approach (Boyd 2019; Heath 2015).

Grouping data points based on their similarity, in this case data points being PSP users, is a commonly used approach in machine learning (Madani 2021). Therefore, even without PSPs' official confirmation, it is likely that PSPs leverage a user's OAI and demographic information to determine what to recommend to other users.

PSPs generate a significant financial value out of personalization, however it is not as measurable as other income streams, such as subscription revenue. Moreover, since these platforms do not offer an alternative service without personalization, a direct comparison of the value generated by their service with or without personalization is practically impossible.

Personalization offers a value circle which increases user's time spent and engagement with the platform, and also generates more data that is used for improving the personalization algorithm itself to retain the user and increase time spent on the platform even more (Belleflamme and Peitz 2019). This value can be indirectly quantified referring to major PSPs' public statements about estimated earnings from personalized services and the contribution of personalized recommendations to the user retention, therefore the subscription revenue (Jannach and Jugovac 2019). Major PSPs publicly announce that a big part of their user engagement comes through personalization. In 2012, Netflix announced through an official blog post that "75% of what people watch is from some sort of recommendation" (Amatriain and Basilico 2012). Similarly in a 2017 interview, Netflix shared that "80 percent of subscribers trust and follow the recommendations of the algorithm" (Chhabra 2017). Through financial lenses, according to a 2015 study conducted by Netflix, personalization saved them $1B annually through user retention and increasing user engagement on the platform (Hunt and Gomez-Uribe 2015) which may likely be a bigger number today given that the platform continued growing its user base since this study. The picture is very similar when it comes to online music streaming giant Spotify. The company offers a "Discover Weekly" playlist for its users, which is a weekly updated list of songs that the platform's recommendation algorithm curates for the user, which has been followed or played by 40 million users in less than a year after its launch in 2016 (Prey 2018). This statistics is not as representative as Netflix's calculations on how personalization enables user retention, however in an official press release in 2020, Spotify addressed Discover Weekly as a "flagship" product and reinforced its contribution to the success of the company (Spotify 2020).

Another contribution of personalization algorithms to PSPs is through original content generation and consumer research. Netflix produces many original series and shows across different countries. They leverage user data as an input to the production process to understand what plots or types of shows would work well with specific audiences (Markman 2019). Similarly, Spotify has been sharing streaming and behavior insights with music research companies as well as artists on their platform as an additional service that their users' data enables (Stassen 2020). Similar to the lack of an exact quantitative measure of how much value personalization brings to PSPs in terms of user retention, these additional benefits of personalization are also not quantifiable but definitely valuable for PSPs.

To sum up, we claim that the majority of the research on users' preferences on OPD is based on free online services and excludes PSPs since they have a different business model than online advertising. However, PSPs benefit from OPD as much as platforms with online advertising and is a venue for extensive research as well.

1.5 Users Benefit from Personalization on Paid Streaming Platforms

After elaborating on how personalization is beneficial to PSPs, now we will explore how beneficial it is to PSP users. We will base our arguments mostly on Budzinski et al.'s comprehensive paper on the subject which examines not only PSPs but also e-commerce and search platforms such as Amazon or Google (2021). This paper uses behavioral economics theories to identify how beneficial personalized search results and recommendations can be for the users of these platforms. The paper starts explaining the most obvious benefits of personalization by referring to other researchers' findings that personalization saves users' time to find the most optimal option in a wide range of available options (Brynjolfsson, Hu, and Smith 2006) and supports this further with empirical evidence that personalization algorithms are more influential on users' buying decisions compared to other online sources of recommendations such as human experts or consumer reviews (Senecal and Nantel 2004).

Unlike social media or e-commerce platforms, there are not many studies conducted on user's benefit specifically from PSP personalization algorithms. Most of the available resources in this subject are limited to small-scale surveys or PSP-led studies. We will give two recent examples conducted on Spotify. According to a 2018 user research conducted by Spotify researchers, only one recommendation that user likes is sufficient to increase user's overall engagement and satisfaction with personalized recommendations (Hosey et al. 2018, p.55-64). An independent research study from 2020 claims that the listening time of playlists created by Spotify with personalized recommendations is equal to half of the listening time of Spotify's general playlists available to all users, such as 'dinner time' or 'road trip' playlists(Mejia 2020). This indicates despite not being the majority, users spend a significant amount of time on Spotify's personalized recommendations. These findings support that personalization is beneficial for a group of PSP users who actively follow personalized recommendations. In the next section, we will discuss whether there are other types of users who would consider personalization as a compromise on ownership of their OPD.

1.6 Why PSP Users May Prefer Opting out of Personalization

We mentioned that PSPs state observing a high user engagement with personalized recommendations, which users may prefer as the most efficient way of finding music or movies of their taste. However, this may not be the only type of PSP user. Budzinski et al. (2003) refers to the rule-following behavior in economics to explain that while some users may prefer efficiency of personalization for routine activities or decisions some others may not consider these as routine activities and prefer their own intuition for similar decisions (Vanberg 2002). From this perspective, we do not know whether listening to music or watching a video is a routine or an outstanding activity for a specific PSP user. For example, you may find it effortless to use Spotify's curated playlist for a casual dinner party, however some users may be quite picky and only listen to the full albums of the artists that they know or recommendations they had from friends in real life.

Finally, and most likely, there are users who are in between. For example, they may occasionally check out the personalized recommendations but are also satisfied with their own content choices or have some classics they always like. Therefore, we can claim that, despite not being the majority, there is a user base in all PSPs who do not use personalized recommendations at all or benefit from them as much as some other users.

Given our assumption that users who do not use personalized recommendations at all would not make the majority, enabling optional personalization may not be a priority for PSPs due to its operational difficulty. However, there may be a broad range of users who may consider the benefits of keeping their OPD completely private and prefer not receiving personalized recommendations, if the option was provided. In the earlier sections, we introduced multiple studies which found that people were willing to pay for the privacy of their OPD, even if the data was not PII. Similarly, the history of songs listened to or movies watched as OAI on PSPs, or data that PSPs leverage for profiling users such as gender or age can be considered as sensitive information for some users. There are two major arguments which may support users' data privacy concerns weighing more than the benefit they receive from personalization, which are the risk of data breach and principle of purposeful data collection.

First argument we will cover is the risk of data breach. PSPs do not share user data with third parties for online advertising, hence a OPD breach may arise from either a cyberattack towards the PSP or a leakage during PSP's own use of data. When it comes to the cyberattacks, according to a study by cybersecurity firm Dynarisk, major PSPs Netflix and Spotify are in top cyberattack risk amongst all websites since they hold users' paid membership and payment information (Dynarisk, n.d.). The type of OPD used for personalization is not directly linked to paid membership or payment data on PSPs, however, it does not eliminate the risk of a cyberattack targeting this type of OPD or leaking it data along with targeted data. When it comes to PSPs' own use of OPD, the amount and detail of OPD these platforms have on users raises concern on the potential extent of this OPD being used beyond personalization. As a one-off example, in 2017, Netflix has posted on their official Twitter account the following "To the 53 people who've watched A Christmas Prince every day for the past 18 days: Who hurt you?". The content of this tweet did not have any malicious intention, was analyzed solely by Netflix based on the data that it legitimately owns and has not revealed any PII about the relevant users. However, it created a public discomfort and even a backlash against Netflix on its ability to micro-target people and track all the actions of its users (Paul 2018). Similar marketing activities in the form of anonymous posts about users' OAI have been used by Spotify as well and the company claimed that they have not received any negative sentiments from their users on these posts (Maheshwari 2017). However, this does not eliminate the risk of such granular data being exposed willingly or unwillingly by PSPs. For example, AOL as an early online web portal in the 1990s had a massive reputation loss in 2006 when they leaked 650,000 users' search queries which released some sensitive information about inappropriate or illegal search terms and partially could be traced back to individuals through some of the terms that have contained PII (Arrington 2006). Therefore, we can claim that the types of songs or movies they searched for or consumed can be very sensitive to some users, which they may not prefer compromising for personalized services.

The second argument we will cover is the principle of purposeful data collection, which is the concept that forms the basis of the European Union's General Data Protection Regulation (GDPR). This concept limits online platforms' OPD collection only to the data that they need for the service and requires the platform to explicitly inform the user about what those OPD are and collect for what purposes (European Union 2013). Especially considering users who do not prefer personalized recommendations, collecting their OPD may conflict with this approach because the information of how their data is utilized by PSPs may not be explicit. We will mention this soon in the next section where we discuss how effective terms and conditions of digital platforms are.

Given that personalization is the default service in PSPs without the ability to opt out, regardless of their choice of using personalization or not, these users will provide their OAI and demographic data to PSP for being utilized in personalization algorithms. Namely, if you have a great taste in music and know some genres very well, you will benefit other PSP users who do not have as much knowledge but interest in that genre, because they may be recommended to listen to other songs you listened to if they have matching songs with your OAI history. The disparity we would like to emphasize here is, the PSP in this example does not have a way to compensate you or differentiate between these two users in terms of cost or the service tier. One user uses the PSP only for access to content they know while the other uses personalized recommendations as well, and the two would be charged the same amount. On a large scale, one type of user would indirectly contribute to the PSP by enhancing their personalization algorithm while not having a personal benefit from it. We claim that PSPs should address this disparity by providing an option to users for opting out of personalization based on their preferences.

1.7 Users' Limited Control of Their Data in Paid Streaming Platforms

A key principle in purposeful data collection is the explicit consent of the user for the specified use of data. Today, this is mostly done through digital platforms providing data processing information in their terms and conditions (TC) or their official website. For example, Spotify TC has a separate section on data privacy which mentions what type of OPD is collected from users and for what purpose. One caveat with their approach is, Spotify does not explain that users' OPD would be utilized for improving their personalization service in general, but rather frames it as "the data collected for providing you a personalized experience" as if user's data is used only for user's own benefit (Spotify Inc. 2021). Netflix as another major PSP does not mention OPD collection and processing for personalization purposes at all in their TC, however provides a section in their official customer support website explaining how a user's OPD is leveraged for recommending content to other users (Netflix, n.d.). A major criticism we can infer from these examples is that PSPs do not explicitly reveal to users how their OPD would benefit the company by improving the company's personalization algorithms. Moreover, even if explained, current TCs are not the most effective way for communicating this data exchange as shown by multiple surveys and studies. For example, a 2017 survey by Deloitte indicates that 91% of people in the U.S. sign up for digital services without reading the TCs (Guynn 2020). An online experiment based on a fictional social media platform finds that users largely do not read or skip available information about data usage and privacy in the TC (Obar and Oeldorf-Hirsch 2018). This may be due to the language, format and the lengths of the TCs or the level of digital literacy of the user being insufficient to comprehend the content of the TC. This paper does not aim to explore the reason behind why TCs are ineffective, but rather focuses on the consequential lack of user awareness on OPD processing since it is the default option PSP users have to opt in for.

Major PSPs currently provide users a certain level of control on their OPD, but it does not reach to a level where users can fully opt in for or opt out of personalization. Spotify is a platform where users can follow each other and if so, view what each other is listening to while both parties are actively using the platform. Users have the option to prevent their followers seeing their activity by starting a private session. However, a user must enable this every time they log back on Spotify, because it resets the user's session preference back to default (Wong 2018). Moreover, private sessions do not prevent Spotify from logging users' activity for the personalization algorithms. In other words, it is not possible to create a session private enough to keep Spotify out (Burgess 2021). Similarly, Netflix enables users to download all their personal data log and reset the data log on the platform, however this only makes Netflix to start collecting OPD from scratch instead of disabling such data collection for good (Collins 2021). When tried for the purposes of this research, it is communicated by Netflix that such an OPD export request may take up to 30 days, which is considerably a long time for a user to access their OPD. Similar to these two major PSPs, none of the PSPs known to us currently provides users a full and certain ability to opt out of personalization. Building on top of the fact that not many users carefully read and evaluate TCs, default and forced personalization may prevent users from being aware of such collection and use of their personal data.

We have mentioned that major PSPs provide a certain amount of information about OPD usage in the extensions of their TC as resources that users can access. However, it still does not eliminate PSPs' responsibility to obtain a data-privacy-first design. As a comprehensive critique of all online service providers, Floridi et al. claim that those services' design forces users to share as much data as possible and therefore it is such services' moral responsibility to safeguard OPD privacy and provide users explicit means of modifying their OPD privacy (2015). Floridi et al.'s review mostly focuses on platforms that leverage online advertising. However, as we explained how PSP benefit from their users' OPD in earlier sections, we can claim that same principle would apply for PSPs as well. We may even argue that a data-privacy-first service design is more critical for PSPs, because user awareness on the extent to which OPDs are leveraged is likely less for PSPs than it is for online advertising platforms. For example, according to a 2019 survey by Pew Research, 79% of Americans claim that they are concerned about companies' usage of their personal data and similarly 80% of them claim that they are concerned about targeted online advertising (Raine et al. 2019). This indicates that users of digital platforms may have more awareness of the data privacy compromise on digital advertising platforms than PSPs. Similarly, another recent survey study on Americans by the University of Oxford finds that the majority of the respondents assume that recommendations by Netflix or Amazon do not use artificial intelligence (Zhang and Dafoe 2019). This may indicate that users associate OPD usage with social media sites, but not necessarily PSPs. They may even assume that since they pay for PSPs, personalization on these platforms is provided by human expert teams rather than algorithms as a premium service. Based on all these, we may claim that PSPs amongst all other digital services have a particular responsibility to inform their users about how their personalization algorithm utilizes users' OPD explicitly.

1.8 Literature Review Summary

Currently, all PSPs offer personalized service by default and they do not offer an option for users to opt out. Referring mainly to Budzinski et al. (2003) and Prince and Wallsten (2020) respectively, we claim that there are PSP users who may not prefer receiving a personalized service or may prefer keeping their data private even if it requires paying for it. On the other hand, the concept of acceptability gap that we introduced with Herzog et al. (2021)'s paper suggests that users may be in favor of personalized services and express data privacy concerns only when it is asked separately. Therefore, we believe that the literature on people's willingness to share data should be expanded with studies that introduce explicit information on how data is utilized as a treatment factor.

We also referred to why PSPs offer personalization as a default service in Section 1.3. Despite the lack of an exact measure, there is strong evidence and statements made by PSPs that users' data is tremendously valuable for PSPs to improve their personalization algorithms and retain users who follow personalized recommendations. Even if a PSP user is not leveraging personalized recommendations, the PSP still benefits from the OPD of that user by improving their personalization algorithm on top of the service fee that user pays. Therefore, a PSP who prevents access to their data would provide less benefit to the PSP than a user who allows it. This brings the question of whether PSPs could charge users different amounts if personalization was offered as an optional service.

Finally, in Section 1.6 we mentioned that PSPs do not explicitly inform users about how users' data is utilized to recommend songs to other users. Given that PSPs are paid platforms and do not show targeted ads like the free internet platforms, there are indications that users of PSPs may assume that their data is not leveraged by the PSP at all. Therefore, PSP users may not be aware of the additional benefit they provide to the PSP by sharing their data. Especially for users who do not benefit from personalization or are sensitive about data privacy, their data may be utilized by the platform without their awareness or a compensation they would consider equivalent.

A user-centric suggestion based on these findings in the literature is PSPs to offer personalization as an optional service. However, there are gaps in the literature to support such claims. Firstly, there are not many studies that measure whether PSP users are aware that on top of paying for the service, they also benefit PSPs by providing their OPD. Secondly, current studies on willingness to pay for data privacy and acceptability gap do not indicate a clear tendency on whether a service without personalization would be perceived as an inferior service or a privacy improvement which may change how much a user is willing to pay for that service. Therefore, the question of whether PSP users would behave differently if they knew about how the recommendation algorithms work and how their willingness to pay for the service would change remains unanswered.

2 Our Study

This paper builds hypotheses based on the unanswered question mentioned in the Section 1.8, specifically for PSPs which are a less researched digital platform type compared to platforms with targeted advertising. Our independent variable is the explicit information of how users' data is leveraged by PSPs to recommend content to other users. Our dependent variable is how much the user is willing to pay for such PSP service. We assume that a user's willingness to pay for the service represents their perception of the value they get from the PSP.

Our first hypothesis (HA) is as follows:

Hypothesis A0 (HA0): If PSP users are explicitly informed that personalization on these platforms are based on their OPD, the amount they would be willing to pay for the PSP would not change.

Hypothesis A1 (HA1): If PSP users are explicitly informed that personalization on these platforms are based on their OPD, they may prefer paying less for the PSP given that they indirectly contribute to the company.

Our assumption here is that, if a group of people are explicitly informed that their data is used by a PSP to recommend songs to others, they may perceive this as a contribution to the platform on top of the service fee they pay. Therefore, compared to people who do not have the explicit information, they may prefer paying a lower fee for the PSP.

Our second hypothesis (HB) is as follows:

Hypothesis B0 (HB0): If PSP users are explicitly informed that personalization on these platforms are based on their OPD and personalization is optional, they would think the fee with personalization would not be different from without personalization.

Hypothesis B1 (HB1): If PSP users are explicitly informed that personalization on these platforms are based on their OPD and personalization is optional, they would think the fee with personalization would be different from without personalization.

Based on the literature introduced so far, we can not infer whether the PSP users would think the service with personalization should be more expensive from the version without personalization. The increased data privacy that comes from opting out of personalization may be something that users are willing to pay extra for. On the contrary, as introduced via the acceptability gap concept, users may not be aware of the trade-off between personalization and sharing data or may be in favor of receiving a personalized service in exchange of data access. Therefore, we will explore whether there is a difference in fee that they are willing to pay for the service with and without personalization.

3 Method

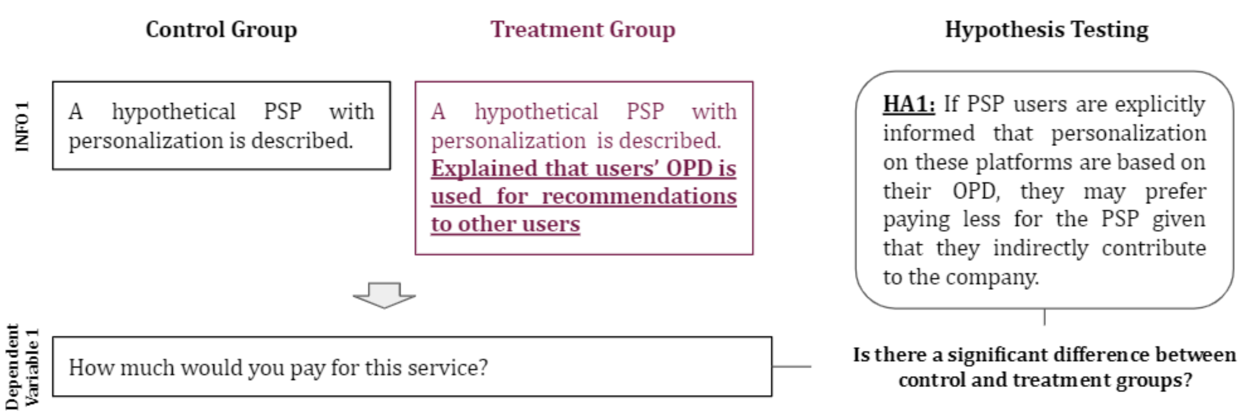
In order to test these hypotheses, a randomized controlled trial (RCT) was designed with one treatment and one control group, all identical except the treatment of providing explicit information on how PSPs leverage OPD for personalization. For convenience of data collection, the RCT was conducted on the online panel platform Prolific, which is widely used for academic surveys and experiments. Starting from here, the variables that are collected from the survey will be referred to by their short names that are available in Appendix A.

The RCT was designed as an online survey in which participants responded to a set of demographic questions and their familiarity with existing PSPs. These were intended to be the control variables of the study. After that, the participants read about a hypothetical PSP providing music streaming services and personalized recommendations to users. With the intention of avoiding priming users' opinion with any existing PSP, a hypothetical platform was preferred. Music streaming as the content type was chosen arbitrarily across different content options that PSPs provide, such as movies or TV shows, and the potential outcomes of this design choice will be addressed later in the Discussion and Further Research section.

Both the treatment and the control groups read a text introducing a hypothetical music streaming PSP which provides access to a large music portfolio and personalized recommendations for the user. The text that the treatment group read was longer with the additional information on how the personalized recommendations work. This information explained that users' OPD would be leveraged by the company to group users based on their similarity and recommend the content that one user listened to a similar user. Details of this text and the treatment factor will be explained in detail in the Survey Design section.

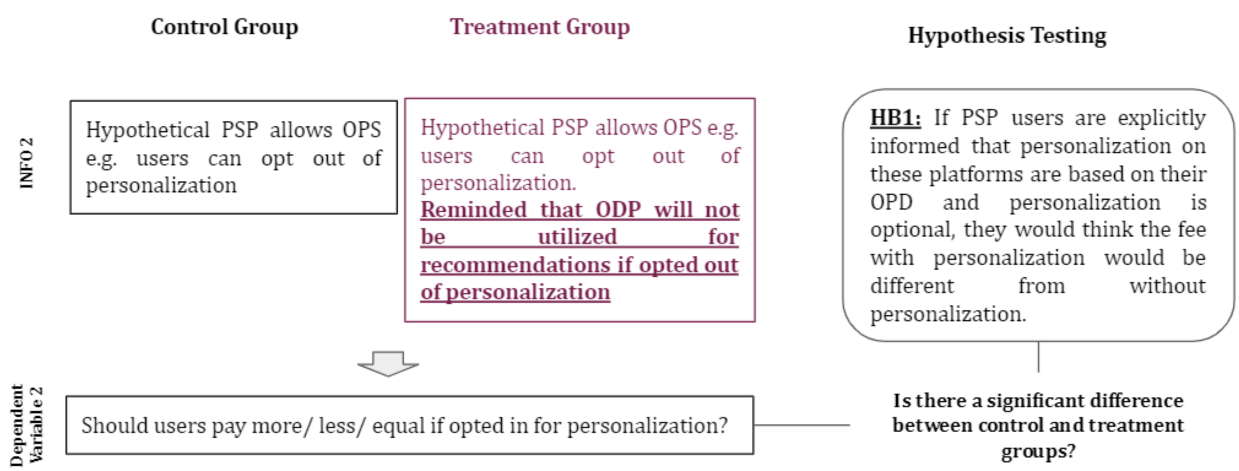
After reading about this hypothetical service, the participants responded to the question that provided the first dependent variable (Y1) of the study which was used for testing HA. Users were asked how much they would be willing to pay for this service in USD. The purpose of this question was to test whether the treatment group would consider the usage of their OPD as an indirect contribution to the service and prefer paying less for the service than the control group does. The demonstration of this hypothesis testing can be seen in Figure 1.

Figure 1: Demonstration of Hypothesis A1



After the first informative text about the PSP and related questions, participants read another text with further information. This text explained that the hypothetical PSP would start providing personalization as an optional service. Respondents read about two different users of the platform, User A who opts in for personalization and User B who opts out of personalization. Both the treatment and the control group read the same information. The treatment group already read how thier OPD was used in the previous section of the survey. At this question, they were reminded that if opted out, the PSP would not leverage users' OPD any longer, to reinforce the intended treatment effect. After reading the text, the participants were asked whether User A and User B should pay the same amount or one should be paying more. Participants were also asked to provide a USD amount for how much each user should pay. The difference between the suggested amount for User A and User B to pay was designed to be used as the second dependent variable (Y2) in order to test whether there was a significant difference between the price that respondents were willing to pay for the PSP service with or without personalization. The demonstration of this hypothesis can be seen in Figure 2.

Figure 2: Demonstration of Hypothesis B1.



The primary statistical measurement method for the hypothesis testing was selected as 2-sample t-test to measure the difference of average Y1 and Y2 across treatment and control groups. Moreover, linear and logistic regression models were also selected to be applied for exploratory analysis on other questions that were asked in the survey to understand underlying preferences of the respondents. The results of these analyses will be discussed in the Results section.

4 Survey Design

The online survey was distributed via Qualtrics. Full survey flow can be found in Appendix B through a link and Appendix C as screenshots of the survey displayed on a mobile device. Survey and question design was based on the principles surfaced in experimental design and survey research sections of Singleton and Straits' comprehensive guide on social science research. (Singleton, Jr. and Straits 2017)

The survey started with demographic questions such as highest level of education completed or political affiliation with the purpose of using them as control factors and observing any unprecedented relationship e.g. income level impacting willingness to pay. After the demographic questions, participants were asked whether they used any or multiple of the following PSPs: Apple Music, Spotify, Netflix, Hulu, Amazon Prime, Disney Plus, HBO Max, Pandora, YouTube Music, Other (free text entry). If participants responded to at least one or more, then they were asked whether they followed and were satisfied with personalized recommendations from these PSPs with the purpose of understanding respondents' prior experience with PSPs which could have an impact on the measured outcomes.

After this set of questions, the respondents read the informative text about a hypothetical PSP as described earlier in the Method section. The treatment factor was introduced at this stage. Both the control and the treatment groups read about a company offering a music streaming service, providing access to music content and personalized song recommendations. The control group read that the company collects their data to recommend them songs. However, they did not read the explanation that their data would be leveraged to recommend songs to other users. Unlike the control group, the treatment group read that the company leverages users' data to find similar users and recommend them the songs that similar users listened which is a method used by existing PSPs as introduced in the Literature Review. They also read that this process would improve the company's recommendation algorithm with the help of user data. The exact texts read in these sections are available in Appendix C.

After the text, there were two understanding checks, one asking whether user's own data would be used for recommendations they receive, and second one asking whether the data was used for recommendations other users received. It was expected for the treatment groups to answer both questions with yes. Control group's answer to the first question was expected to be yes, but the second answer would vary based on their prior knowledge i.e. they already knew how PSP personalization algorithms work.

After the understanding checks, the participants responded to three Likert scale questions which measured how much respondents think that themselves as a user, other users and the company benefits from respondent's data on this platform. These questions aimed to measure the effectiveness of the treatment factor by measuring the difference between control and treatment groups' perception on how PSPs benefit from users' OPD. Outcomes of these questions are going to be analyzed in Further Analysis subsection of the Results section.

After these questions, the respondents were asked about their willingness to pay for this hypothetical PSP in USD, which created Y1, the first main dependent variable of the study. A range was provided from $1 to $20 with $1 increment, taking Spotify's monthly fee for different tiers from $4.99 to $15.99 as a real-life benchmark. The main disadvantage of providing such a range was anchoring participants, however such design was preferred in order to avoid outlier answers and the risk of anchoring was tried to be mitigated by providing a larger range than the average price of an existing PSP.

The next part of the survey introduced a new informative text. Both groups read that the hypothetical PSP would now offer personalization as an optional service. Two users were introduced to the respondents, User A opting in for personalization and User B opting out. Unlike the control group, the treatment group read additional informaiton about the PSP leveraging only User A's data for recommendation algorithms in order to remind the treatment factor. The exact texts are available in Appendix C.

The respondents were asked three Likert scale questions which measured the importance of personalization being optional for the user and whether they would opt in or not. Similar to Likert scale questions after the first informative text, these questions intended to measure the effectiveness of the treatment factor and the outcomes of these questions are going to be analyzed in Further Analysis under the Results section.

The final part of the survey asked how much User A and User B should pay for the service, and the difference between the two values were calculated to create Y2, the second main dependent variable. The respondents were able to declare that User A and User B should pay equally, which aimed to capture their true opinion but also limited the marginal difference between User A and User B. This will be addressed as a limitation in the Discussion and Further Research section.

5 Sample Size and Data collection

The study sample was intended to cover existing or potential users of PSPs in order to capture an audience that can comprehend and compare a trade off between a personalized streaming service and data sharing. Given that my study referred to a hypothetical music streaming service, the age range was selected as 18 to 45, inclusive, in order to focus on the user range that makes the majority of such services according to a public market research about Spotify as a real life music PSP, based on publicly available data on Statista (Hlebowitsh 2021). Moreover, given that the availability of PSPs differ across the world and the survey aimed to collect information on which platforms respondents had experience with as a control factor, the sample was also limited to respondents residing in the United States. Apart from age and location, the sample was intended to be equally distributed between males and females and no other demographic specifications were aimed.

Sample size of the experiment was determined by using a standard power analysis calculator provided publicly by University of British Columbia, which is based on the expected mean value of the dependent variable for each group and the expected standard deviation (Brant, n.d.). The first assumption was that the average monthly fee that the control and treatment group would prefer paying would be \$5 and \$6 on average, with a standard deviation of \$4. These assumptions were centered on the range from Spotify's monthly fee \$9.99 as a real life benchmark which was expected to be the upper bound and \$1 as the lowest non-zero fee a respondent could answer. The unit for responding to this question was set to \$1 for simplicity. Based on these assumptions, a sample size for 80% power rate at 95% confidence interval was 252 respondents for each group and 504 in total. Concerns and limitations for sample size will be discussed in Discussion and Further Research section.

Respondents were recruited through the Prolific online panel. They received a compensation that was suggested by Prolific for expected 5-8 minutes of survey time and also read a short information blurb about the survey. Full survey materials are shared in Appendix B and Appendix C. In case of a discrepancy that can be found later in data, a 5% buffer was planned and 530 responses were collected in total. Participants who failed the attention check were automatically rejected. Out of 530 responses collected, 3 of them indicated that the text that conveyed the information about the hypothetical PSP were not understood, based on the understanding check questions. These participants were still compensated, however excluded from the study given that they may have not received the information as expected in the research design. The final sample size was N=527 people, 263 from control and 264 from treatment groups.

The data was collected on the Qualtrics platform and exported as a CSV file. The data file that was used for the main analysis was prepared manually using Microsoft Office Excel functionalities. Appendix B provides public links to the Github repository that includes the raw data file, the raw data file excluding 3 participants failing the understanding check and the processed data file used for analysis. The Github repository also provides a Readme that explains the steps taken for the processed data file. Appendix A provides a table that introduces which variable corresponds to which data field and how they will be referred to in the Results section.

6 Sample Demographics

The online sample setting aimed at a 50%-50% gender distribution across males and females, however the gender identification in the results will be based on users' self-identification. According to the self-report gender question in the survey, control group's 49% was male, 46% was female, 4% was non-binary and 1% preferred not to say. Treatment group's 49% was male, 48% was female, 3% was non-binary without any respondent preferred not to say. In the age range set between 18 to 45, the median age was 27 and 28.5 while the average age was 28.2 and 29.4 for control and treatment groups respectively. Most of the participants completed high school or college at a minimum, representing 86% of all the sample, almost equally distributed across control and treatment. No participant responded not attending school or completing only elementary school as their highest level of education that was completed. Majority of the sample had a net monthly average income of $2000-$4999 with 35% and <$1999 with 33% respectively. Treatment group had a slightly bigger proportion of higher income than the control group. $5000-$9999 income level made up 19% of the treatment group while it made only 15% of the control group and <$1999 made up 31% of the treatment group while it made up 35% of the control group. According to the self-report political view question in the survey, 54% of the sample was democrat while the second biggest proportion was independent, with 28%. Only 12% of the total sample was republican. The proportion of the independents was higher for the treatment group, which was 31% compared to 25% in the control group. Summary of demographics are also available in Table 1.

Table 1: Sample Demographics

				Proportions		
	Control	Treatment	Total	Control	Treatment	Total
Average Age	28.2	29.4	28.8	-	-	-
Median Age	27	28.5	28	-	-	-
Gender						
Female	121	127	248	46%	48%	47%
Male	129	130	259	49%	49%	49%
Non-binary	11	7	18	4%	3%	3%
Prefer not to say	2	0	2	1%	0%	0%
Education						
Graduate school	35	36	71	13%	13%	13%
College/ university	139	141	278	53%	53%	53%
High School	87	90	175	33%	34%	33%
Middle School	2	1	3	1%	0%	1%
Monthly Average Income						
$20000 or more	10	17	27	4%	6%	5%
$10000 to $19999	14	15	29	5%	6%	6%

$5000 to $9999	39	50	89	15%	19%	17%
$2000 to $4999	97	89	186	37%	34%	35%
Less than $1999	91	81	172	35%	31%	33%
Prefer not to say	12	12	24	5%	5%	5%
Political View						
Democrat	148	135	283	56%	51%	54%
Republican	34	31	65	13%	12%	12%
Independent	66	83	149	25%	31%	28%
Something else	13	11	24	5%	4%	5%
Prefer not to say	2	4	6	1%	2%	1%
TOTAL	263	264	527	50%	50%	100%

Second set of demographic questions investigated participants' paid streaming platform usage. Amongst all participants, 11 of them reported not using any paid streaming platform (2%), 45 of them (9%) reported using only one paid streaming platform and 471 of them (89%) reported using more than one. As seen in Figure 3, amongst the participants used only one platform, Netflix, Spotify and Amazon Prime / Amazon Music were the top three platforms used. As seen on Figure 4, amongst all participants used one or more platforms, top 3 were Netflix, Amazon Prime / Amazon Music and Hulu.

Figure 3: Participants using only one paid streaming platform

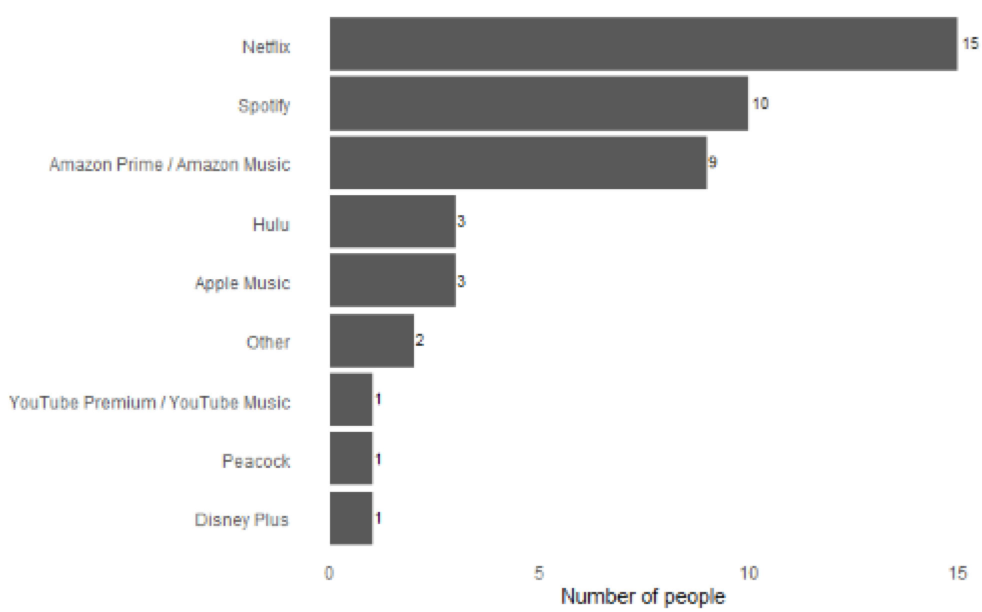
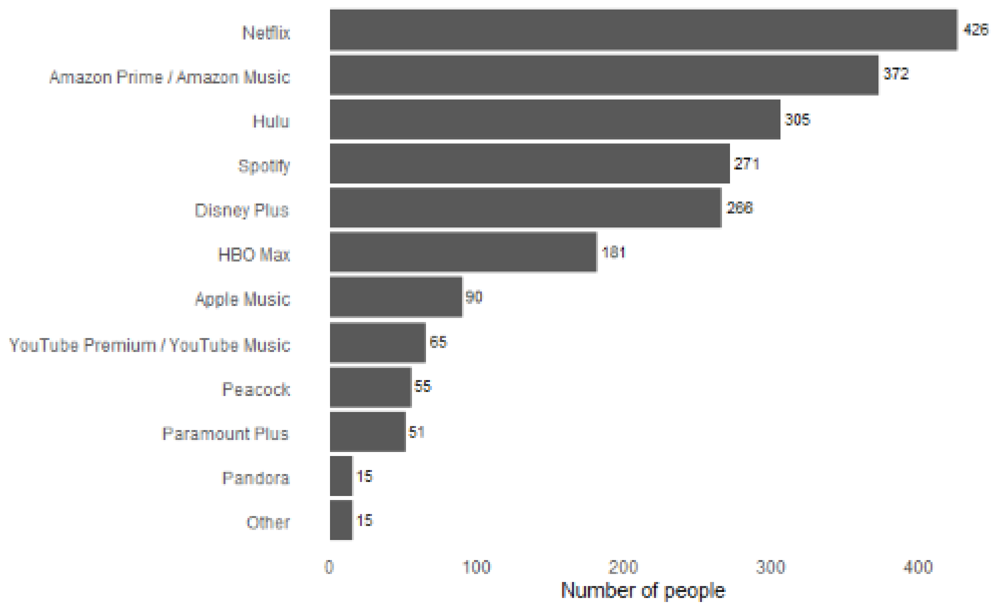


Figure 4: Participants using at least one or more paid streaming platform



516 participants who used at least one paid streaming platform were asked how much they follow or consume the content recommended to them by personalization algorithms on paid streaming platforms, on a 1-5 Likert scale of never, rarely, sometimes, most of the time, always. The majority reported following personalized recommendations rarely (22%) or sometimes (58%), which was distributed similarly for both the control and treatment groups. 35 participants (7%) reported never using or following personalized recommendations. These results can also be seen in Table 2.

Table 2: Preferences for Personalization on Existing Paid Streaming Platforms

				Proportions		
Following Personalized Content	Control	Treatment	Total	Control	Treatment	Total
Always	5	6	11	2%	2%	2%
Most of the time	35	23	58	14%	9%	11%
Sometimes	146	153	299	57%	59%	58%
Rarely	54	59	113	21%	23%	22%
Never	16	19	35	6%	7%	7%
TOTAL	256	260	516	100%	100%	100%

The respondents who reported following personalized recommendations rarely or more were asked a follow-up question to evaluate their satisfaction with the personalized recommendations on these platforms. The majority reported being satisfied with the recommendations sometimes (54%) or most of the time (31%). Only 8 respondents (2%) reported that despite following personalized recommendations, they were never satisfied with them. These results can also be seen in Table 3.

Table 3: Satisfaction with Personalization on Existing Paid Streaming Platforms

				Proportions		
Personalized Content Satisfaction	Control	Treatment	Total	Control	Treatment	Total
Always	4	2	6	2%	1%	1%
Most of the time	67	82	149	28%	34%	31%
Sometimes	142	119	261	59%	49%	54%
Rarely	25	32	57	10%	13%	12%
Never	2	6	8	1%	2%	2%
TOTAL	240	241	481	1	1	1

7 Results

All results will refer to the short names of the variables introduced in Appendix A e.g. Dependent Variable 1 will be called Y1. Below is the recap of hypotheses which explained in detail in Our Study section previously.

HA0 : There is no significant difference between how much the treatment and the control groups would be willing to pay for the monthly PSP fee.

HA1: There is a significant difference between how much the treatment and the control groups would be willing to pay for the monthly PSP fee.

HB0: There is no significant difference between how much the treatment and the control group would be willing to pay for the PSP service with personalization versus without personalization.

HB1: There is a significant difference between how much the treatment and the control group would be willing to pay for the PSP service with personalization versus without personalization.

7.1 Hypothesis Tests

In order to test the HA, we used the 2-sample t-test as a commonly preferred method to compare the means of two different groups, which is the average amount of service fee, Y1. The 2-sample t-test has two assumptions (STHDA, n.d.). First one is that the data should be normally distributed, which can be waived if both samples have at least 30 data points, based on the Central Limit Theorem (LaMorte 2016). Second one is that the two groups should have equal variances.

We checked the normality of Y1 using the Shapiro-Wilk Test and found that Y1 is not normally distributed ($W=0.93427$, p-value ≤ 0.01). We checked the variances by using F-test and found that the variances of the two groups are not significantly different ($F_{262,263}=0.85$, $p=0.2$).

Given that the sample size was sufficiently large and the variances were equal, we used the 2-sample t-test to compare the mean Y1 across treatment and control groups. The 263 participants in the control group (M=8.05, SD=2.9) compared to the 264 participants in the treatment group (M=8.4, SD=3.1) did not report a significantly different monthly fee for the PSP, t(525)=0.28198, p=.0.78. We rejected HA1 and accepted the null hypothesis, which claims that the explicit information on how PSPs leverage a user's OPD to recommend songs to other users does not significantly change how much the user prefers paying for the platform.

In order to test HB, we needed to explore whether there was any significant difference between the control and treatment groups on Y2, which was the difference of preferred price of the hypothetical PSP with personalization and without personalization. Y2 did not follow the normal distribution as well (W=0.67647, p-value$\leq$0.01) and had almost different variances across the two groups ($F_{262,263} = 0.8$, p=0.08). Therefore, we applied both the 2 sample t-test and the Welch's t-test for Y2, since the latter does not have the equal variances assumption (Lakens 2015).

Both tests resulted very similarly. According to Welch's t-test, the 263 participants in the control group (M=0.81, SD=1.9) compared to the 264 participants in the treatment group (M=0.5, SD=2.2) did not report a significantly different Y2, t(525)=0.28198, p=.0.78. We rejected HB1 and accepted that there was no significant difference between how much the treatment and the control groups are willing to pay for the PSP service with personalization versus without personalization.

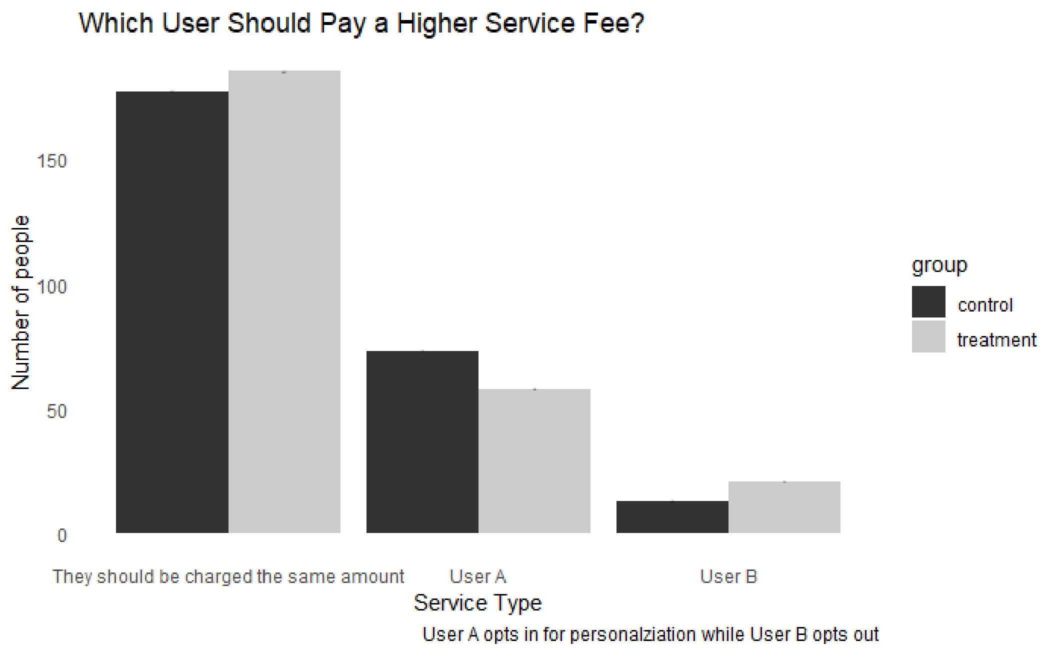
Table 4: Responses to Y3, Y4 and deducting Y2 as the difference of the two

	Fee with Personalization	Fee without Personalization	Difference
Control	\$8.41	\$7.60	\$0.82
Treatment	\$8.21	\$7.67	\$0.54

Given that Y2 was derived from the difference of Y3 and Y4, how much users would pay the PSP with personalization versus without personalization respectively, we looked further into respondents' input for each case as seen in Table 4. As seen in Figure 5, the majority of the respondents reported that the service should be the same price regardless of personalization being provided or not. Amongst the ones who reported a difference, the majority of both the control and treatment groups reported that the fee with personalization should be higher. This indicates that regardless of the treatment factor, respondents mostly consider personalization as a service to be paid for.

As a supplementary analysis, we applied Welch's t-test for both Y3 (t(516.13)=0.53,p=0.6) and Y4 (t(521.88)=-0.45,p=0.6),and found no statistically significant difference. Normality and variance tests for these variables are available in Appendix D.

Figure 5: Participants' responses to whether User A and B should pay different fees



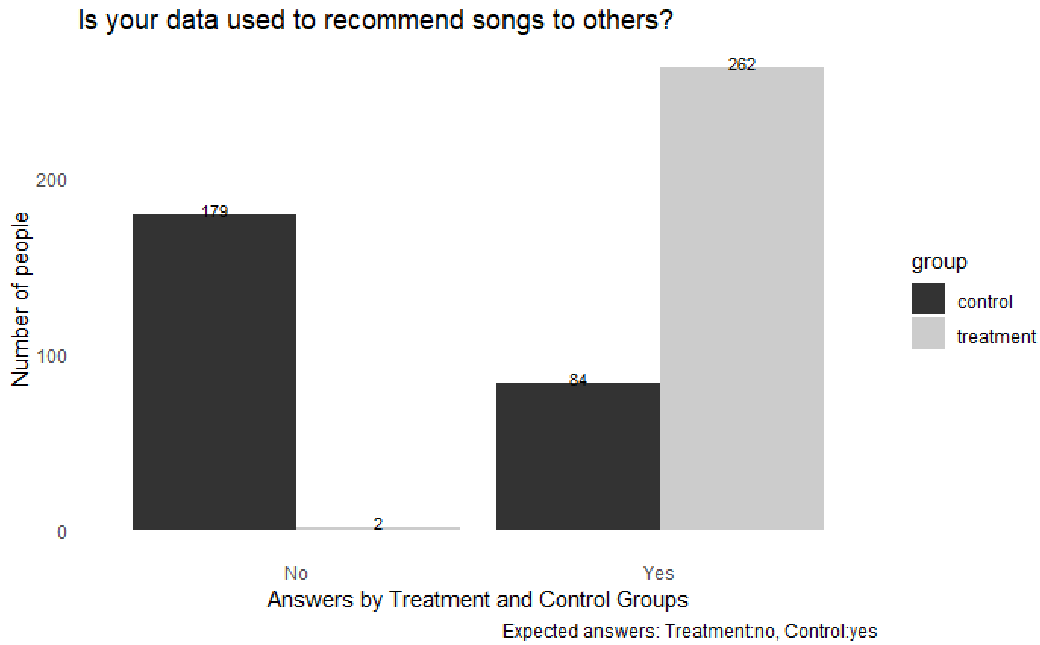
7.2 A Different Approach to Experiment Grouping

The main difference between the treatment and the control groups in our study was the explicit information on how users' data is used in personalized recommendations. However, despite not being exposed to in the experiment, the control group could still have this information based on their prior knowledge about PSPs and demonstrate a similar tendency that was intended for the treatment group. As shared earlier, there was an understanding check that asked "whether your data is used for recommending songs to others" which treatment group was expected to answer with "yes" and control group was expected to answer with "no". Figure 6 shows that the majority (68%) of the control group responded "no" to this question as expected, however, 32% of them still responded "yes" which indicates that they may have behaved similarly to the treatment group to some extent.

In order to explore this, we have ignored the treatment and control groups and regrouped all respondents based on the answer they gave to the treatment understanding check. Amongst all participants, 346 of them responded "yes" to the intended treatment effect while 181 of them responded "no", which made the reconstructed treatment and control groups respectively. For the simplicity of this heuristic approach, we did not explore the other demographic factors that were counted in random distribution of the respondents. We will still refer to the hypotheses tests with the original groups as the main results of the study.

Applying same variance tests and assuming sufficiently enough large sample, 2-sample t-test did

Figure 6: Respondents' answers to the understanding check of the treatment factor



not find any difference on Y1 for reconstructed groups. For Y2, given that the variances of the two groups were almost different ($F_{262,\ 263}$=0.8, p=0.08) we applied both 2-sample t-test and Welch's t-test. Both tests resulted similarly. According to Welch's t-test, Y2 was significantly different between the reconstructed treatment(M=0.46,SD=2.07) and control(M=1.09,SD=2.11) groups, (t(360.52)=3.24, $p \leq 0.05$). Therefore, if we assume that reconstructed groups are a better representation of the treatment factor, we accept HB1 and claim that the two groups were willing to pay significantly different amounts for the PSP with personalization versus without personalization.

Since Y2 is constructed from Y3 and Y4, we look into each using Welch's t-test. Between the reconstructed control(M=7.64,SD=2.87) and treatment (M=7.66, SD=3.1) groups, there is no significant difference on Y4(t(391.02)=-0.08,p=0.9). However, reconstructed treatment (M=8.73,SD=3.01) and control (M=8.13, SD=3.26) groups reported significantly different amounts for Y3v(t(392)= 2.11, p=0.03). This means that for the PSP service with personalization, the reconstructed treatment group is willing to pay significantly less than the reconstructed treatment group. A summary of these results can also be seen in Table 5.

Table 5: Y2, Y3 and Y4 based on reconstructed experiment groups

Reconstructed Groups	Fee with Personalization	Fee without Personalization	Difference
Control	$8.68	$7.59	$1.08
Treatment	$8.13	$7.66	$0.47

7.3 Further Analysis

Apart from the main dependent variables collected for the purpose of hypotheses testing, we analyzed other data in our survey which could serve as dependent variables and provide further insights into people's perceptions towards personalization in PSPs. We ran 10 different multiple linear regressions, one for each dependent variable, by including rest of the variables as independent variables in the model. The aim of this heuristic exploratory approach was to detect potential correlations between any two variables, controlled by other factors. Given that our main goal was exploratory but not predictive analysis, we mainly referred to R-squared for the strength of the model in explaining the variance in the dependent variable.

Despite the sample size of N=527 is considered sufficient for the normality assumption of linear regressions, we also built logistic regression models. We did this by creating a binary version for each numeric dependent variable, which resulted in 0 if the value was smaller than the average of the dependent variable, 1 if equal to or greater than the average. Out of 10 models, we selected 4 that were consistent in terms of statistical significance across linear and logistic regression outputs. We used ANOVA to simplify each model down to significant variables only. For example, Model1 started with 15 independent variables including all variables in the data set, and narrowed down to only 2 independent variables that were statistically significant and introduced in this paper. The results of the final models are displayed in Table 6 and Table 7.

For consistency with the data set, the variable names in the models are kept as originals and results are summarized using the abbreviations of each variables, available in Appendix A. The complete analysis, including the models not selected for this paper, the larger versions of the simplified models and the ANOVA analysis are available in the code shared in Appendix D.

7.3.1 Model1: Users are willing to pay less if their data benefits others

This regression model is applied on Y1, the main dependent variable which did not significantly change across control and treatment groups. However, both the linear and logistic regressions models pointed out that Y6 was statistically significant on Y1. Y6 was how much the respondents agreed with the statement of "Other users benefit from my data on this platform a lot." on a scale from 1 to 5. For each unit of more agreement, respondents were willing to pay $0.25 less on average for the monthly fee of the service, controlled for all other factors ($p<0.05$). This finding indirectly supports HA1 by indicating that if a user is aware of how their data is used for the benefit of other users, they may consider this as a contribution towards the service and prefer paying less for it. However, given that the Multiple R-Squared can explain only 3.8% of the variance of Y1, we still can not claim a significant relationship between the willingness to pay for the service and awareness of how one's data is used for benefit of others.

Table 6: Regression Models on Y1 and Y5

	Dependent variable:			
	Y1	Binary Y1	Y5	Binary Y5
	OLS	*logistic*	*OLS*	*logistic*
	(1)	(2)	(3)	(4)
Constant	10.010***	0.898*	1.647***	−4.143***
	(0.683)	(0.476)	(0.143)	(0.565)
benefit_4	−0.256**	−0.118	0.208***	0.560***
	(0.118)	(0.082)	(0.034)	(0.120)
Income: Less than $1999	−1.218**	−0.618		
	(0.598)	(0.415)		
Income: $2000 to $4999	−0.383	−0.122		
	(0.594)	(0.413)		
Income: $5000 to $9999	−1.291**	−0.687		
	(0.635)	(0.440)		
Income: $20000 or more	0.537	0.205		
	(0.793)	(0.561)		
Income:Prefer not to say	−0.703	−0.050		
	(0.821)	(0.573)		
Group:treatment			−0.238***	−0.892***
			(0.074)	(0.260)
Gender: Male			−0.196***	−0.558**
			(0.066)	(0.219)
Gender: Non-binary / third gender			−0.355**	−1.080*
			(0.178)	(0.636)
Gender: Prefer not to say			−0.163	0.377
			(0.515)	(1.580)
Personalization_satisfaction			0.267***	0.625***
			(0.030)	(0.111)
Political: Independent			−0.054	0.007
			(0.074)	(0.248)
Political: Prefer not to say			−0.619**	−0.607
			(0.299)	(1.033)
Political: Republican			−0.055	0.084
			(0.098)	(0.336)
Political: Something else			−0.115	−0.521
			(0.154)	(0.538)
preferences_5			0.206***	0.540***
			(0.027)	(0.093)
Observations	527	527	527	527
R^2	0.038		0.392	
Adjusted R^2	0.027		0.379	
Akaike Inf. Crit.		727.605		573.479
Note:				*p<0.1; **p<0.05; ***p<0.01

7.3.2 Model2: Treatment group does not benefit from their data on the platform a lot

This model on Y5 looks at how much the respondents agree with the statement of "As a user, I benefit from my data on this platform a lot." from 1 to 5. The treatment group agreed with this statement 0.238 points less than the control group did on average, controlled by other factors ($p < 0.01$). The difference is not as big as 1 point on a scale from 1 to 5, but it still indicates that if a user is informed explicitly about how their data is used for others' benefit on a platform, they may think that they don't benefit from their own data on this platform as much.

Two other findings of this model are linked to each other. More someone stated that they were satisfied with personal recommendations, variable C14, and the more someone agreed with the statement of "I would opt in for personalization", variable Y10, the more they agreed with the statement in Y5 ($p < 0.01$). This finding supports the literature review that there is a group of PSP users who are satisfied with personalization services and not as concerned about how their OPD is used as some other users may be.

The significance of Y6 on Y5 is expected since both of the experiment groups reported that they believe that the PSP leveraged their OPD for other users. The model also suggests that gender has a significant impact, which was not expected given that there was no assumption on gender difference. The multiple R-square of this model suggests that these variables explain 39% of the variance in Y3, which indicates that it is a considerably strong model.

7.3.3 Model3: Optional personalization is preferred if company benefits from users

The next model is on Y8, which is how much the respondents agree with the statement of "The company should make the personalization optional." from 1 to 5. Y5 is found statistically significant. For each point more someones agrees with the statement of "the company benefits from my data a lot" from 1 to 5, they agree with the personalization being optional by 0.46 points more on average, controlled by other factors ($p < 0.01$). This indirectly supports our assumption in the literature review that some users who do not benefit from personalization as much as other users may prefer personalization being optional.

Other two significant variables are Y9 and Y10, which stand for how much the respondents agreed with the following statements respectively: "It doesn't matter a lot if the personalization is optional.", "I would opt in for personalization.". As expected, these two are inversely correlated with Y8 ($p < 0.01$). This model has a multiple R-squared of 23.5%, which makes it considerably strong in explaining the variance of Y8.

Table 7: Regression Models on Y8 and Y10

	Dependent variable:			
	Y8	Binary Y8	Y10	Binary Y10
	OLS	*logistic*	*OLS*	*logistic*
	(1)	(2)	(3)	(4)
Constant	4.770***	0.942	2.056***	−1.811**
	(0.190)	(0.603)	(0.388)	(0.877)
preferences_4	−0.294***	−0.818***	0.104**	0.127
	(0.030)	(0.107)	(0.047)	(0.104)
preferences_5	−0.107***	−0.308***		
	(0.026)	(0.083)		
benefit_5	0.138***	0.441***		
	(0.036)	(0.117)		
preferences_1			−0.232***	−0.475***
			(0.063)	(0.146)
benefit_1			0.462***	0.878***
			(0.059)	(0.138)
Personalization_satisfaction			0.167***	0.260**
			(0.047)	(0.103)
Observations	527	527	527	527
R^2	0.235		0.267	
Adjusted R^2	0.231		0.262	
Akaike Inf. Crit.		619.273		615.483
Note:				*p<0.1; **p<0.05; ***p<0.01

7.3.4 Model4: Users satisfied with personalization do not care about optional personalization

The last model is on Y10, which is how much the respondents agree with the statement of "I would opt in for personalization." from 1 to 5. As expected, C14 is statistically significant on Y10, which suggests that more satisfied someone with personalized recommendations, more willing they would be to opt in for personalization ($p < 0.05$). Another significant variable Y8 suggests that the more someone thinks that personalization should be optional, less they would be willing to opt in for personalization. The multiple R-square of this model is also considerably high at 26.7%.

8 Limitations

We will introduce four main limitations our study had. Firstly, our survey may have introduced the treatment factor so implicitly that the intended impact may have not been achieved. Moreover, respondents were not forced to state a difference for the questions where the difference was measured as the dependent variable. However, these design choices were made specifically in order not to prime users for data privacy concerns, which we believe may be the case for existing literature on data privacy. Second limitation was the sample size. Given that there was no similar study as a benchmark, our assumption for the dependent variable values used for power analysis were smaller than the actual findings. We also limited the sample size in a feasible range that could be covered with the research fund available for this master thesis. If calculated with actual findings, the sample size of the study should have been a lot higher, which is recommended for any replication or adaptation of this study. Third limitation is that the sample being recruited through an online panel, which is not representative of the total population. The fourth limitation is that the hypothetical service in the survey being a music streaming service may not be representative of all PSPs, hence the results could have been biased by different perceptions on existing similar services e.g. Spotify. However, respondent's familiarity with PSPs or specifically Spotify were not found to be statistically significant factors in any of our analyses, but the treatment effect may have been different if the survey introduced a different PSP.

9 Discussion and Further Research

Our main hypothesis tests suggest that specifically for PSPs, people do not consider personalization algorithms as a data privacy compromise or a contribution that they are not compensated fairly for. Moreover, if PSPs were to make personalization optional, the majority of our sample do not consider opting out of personalization as an additional data privacy benefit and believe that the service should cost less without personalization. Despite our main hypothesis, we also have slightly different indications from further analysis of the data collected. Our results indicate that if people have the perception that their data is being used for others, which is the case in how PSP recommendation algorithms work, then they are willing to pay less for the PSP. We also found that if people believe that PSP companies benefit from their data a lot, they are also more in favor of personalization being optional on these platforms. In the literature review, we suggested that a group of PSP users may not be using personalized recommendations as much as other users do. Our study did not find evidence that these users may prefer opting out of personalization because of data privacy concerns, but they may still prefer opting out and also paying less for the service. However, given that our results do not indicate that users consider personalization being default on PSPs as an unfair contribution to the platform or a data privacy concern, then there is not much incentive for PSPs to offer personalization as an optional service given that they attribute a big part of their value creation to their personalized services.

Another concept we introduced in the literature review was the acceptability gap, which addresses the duality between people's reported concern on data privacy and willingness to receive personalized services online. Our findings support the latter. Explicit information on their OPD being utilized by the PSP to recommend songs to other users did not create a significant change on how much a user would be willing to pay for such service. One may argue that this is because the type of data PSPs use, such as users' music taste or songs they listen to, is not sensitive. However, the literature review on data privacy concerns show that people attribute a certain data privacy concern and even willingness to pay for any type of personal data. Therefore, we claim that when people are explicitly asked for their concern on data privacy, they may be primed to state that they are concerned. When their concern is attempted to be measured implicitly by introducing the services they receive in exchange of personal data, they are not primed as such and do not indicate as much concern as they would when asked explicitly.

As we introduced the limitations, this study was limited to a single type of PSP and a small sample that was calculated for the assumption of a bigger difference in dependent variables. Given that further findings and A Different Approach to Experiment Grouping section found indirect support for our hypotheses, we recommend further research with a similar concept but a few key changes. For example, a further study could include six groups, half reading a movie and the other half reading a music streaming service and adjusting the wording of how company benefits from OPD in a gradually increasing way for three groups under each PSP. From a much bigger sample for an expected average difference of \$0.1 to \$0.5 between groups. These groups should also have a bigger sample size for the small margin of difference. Such study would eliminate a big part of the limitations of our study and produce a more powerful result.

This study was designed for all existing and potential users of PSP as the main population of the study, however there was no feasible way to limit the sample to this subset assuring randomization. However, users of the existing platforms are much likely to give true opinions on the compromise and satisfaction between data sharing and receiving personalization. We encourage existing PSPs to consider optional personalization in their user research agenda which could lead to a more powerful study in which our intended treatment factor would be replaced by the real life experience.

A final note on further research is based on the location of respondents. As shared in the Literature Review, different countries have different legislation on OPD processing. This may lead to different levels of concern and preferences for PSP as well. For example, further research could compare a US sample to an EU sample, in which the latter would be expected to have more concerns on data sharing given the stricter laws on personal data in the EU.

10 Conclusion

Online personal data becomes more of an asset for many digital businesses and there are conflicting findings about people's perception of their data online. Referred to as "acceptability gap" in the literature, people both tend to express data privacy concern and even may be willing to pay for the privacy of their online data regardless of how sensitive that data is. On the other hand, people are also happy to receive services in exchange for their data, such as personalized recommendations. Our study indicates that for paid streaming platforms, user attitude is likely towards the latter. We expected that people who already pay for these platforms may consider their personal data as an additional benefit to these companies or preserving their data as an additional data privacy service and assume their willingness to pay for the service would change if they knew how their data is used. Hence, we claimed that these platforms should enable optional personalization. However, we found that people's willingness to pay for the service does not change as such. They consider personalization as an additional service which could be paid extra for, if it was ever optional. We found indications that knowing how much these companies benefit from their data could lead people to be willing to pay less, and recommend further research on that. Our research adds on literature about online personal data, showing that asking people directly about concerns on control of their personal data and privacy may not reflect their true opinion or behavior in day to day life. On the other hand, we also find evidence that there is potentially a user group of paid streaming platforms who do not benefit personalization as much as other users and could prefer receiving the content service without personalization. Based on our results, we do not claim that optional personalization in paid streaming platforms is a desired feature by potential users. However, we still find indications that users' awareness of how much their data is utilized by these platforms is limited and further research can reveal findings that support our hypotheses, especially conducted directly by paid streaming platforms on existing users and more explicit measurement of the value that paid streaming platforms gain from user data.

11 Appendices

Appendix A

Table of Variable Abbreviations:

Variable Name	Description	Variable Type	Short Name	Raw Data / Deducted
Age	Age in 2021	Control	C1	Deducted
Gender	Reported gender identification	Control	C2	Raw
Education	Highest level of education completed	Control	C3	Raw
Education_numeric	Highest level of education recoded from 4 to 1, highest to lowest available in data	Control	C4	Deducted
Income	Monthly average net income	Control	C5	Raw
Political	Political affiliation	Control	C7	Raw
Platforms	Which PSPs respondents use, chosen from a list	Control	C8	Raw
Platforms_text	Which PSPs respondents use, "other" entry as free text	Control	C9	Raw

Using _PSPs	Number of platforms used, recoded as 2: more than one platform, 1: only one platform	Control	C10	Deducted
Using_Spotify	Binary variable, whether respondent uses / has used Spotify, 1:yes, 0:no	Control	C11	Deducted
Personalization	Whether respondent follow personalized recommendations on the platforms they use	Control	C12	Raw
Satisfaction	Whether respondent is satisfied with the personalized recommendations on the platforms they use	Control	C13	Raw
Personalization_satisfaction	Whether respondent is satisfied with the personalized recommendations on the platforms they use, recoded as: Never:1, Rarely: 2, Sometimes:3, Most of the time:4, Always:5	Control	C14	Raw
fee	Monthly fee respondents willing to pay for the service	Dependent	Y1	Raw
Personalization_fee_difference	Subtraction of fee_2_2 from fee_2_1	Dependent	Y2	Raw
Fee_personalization	Fee with personalization	Dependent	Y3	Raw
Fee_without_personalization	Fee without personalization	Dependent	Y4	Raw
benefit_1	1 to 5 Likert scale on agreement with the following: As a user, I benefit from my data on this platform a lot.	Dependent	Y5	Raw
benefit_4	1 to 5 Likert scale on agreement with the following: Other users benefit from my data on this platform a lot.	Dependent	Y6	Raw
benefit_5	1 to 5 Likert scale on agreement with the following: The company benefits from my data on this platform a lot.	Dependent	Y7	Raw
preferences_1	1 to 5 Likert scale on agreement with the following: The company should make the personalization optional.	Dependent	Y8	Raw
preferences_4	1 to 5 Likert scale on agreement with the following: It doesn't matter a lot if the personalization is optional	Dependent	Y9	Raw
preferences_5	1 to 5 Likert scale on agreement with the following: I would opt in for personalization.	Dependent	Y10	Raw
equal_fee	Whether respondents think the fee with or without personalization should be the same	Dependent	Y11	Raw
group	Treatment / control group assignment	Independent	I1	Deducted
Duration (in seconds)	Time spent by survey respondent	Not used	N/A	Raw
prolific_ID	Survey respondents' unique Prolific ID	Not used	N/A	Raw
Date_of_birth	Date of birth	Not used	N/A	Raw
understand_you	Understanding check: does platform use your data to recommend songs to you, expected answer is "yes" for both groups	Understanding check	N/A	Raw
understand_other	Understanding check: does platform use your data to recommend songs to others, expected answer is "yes" for treatment, "either yes / no" for control	Understanding check	N/A	Raw

Appendix B

Data files are shared in the public Github repository, Link to detailed survey questions, survey flow and contact details for detailed questions are avaialble in the Readme of the Github repository.

Appendix C

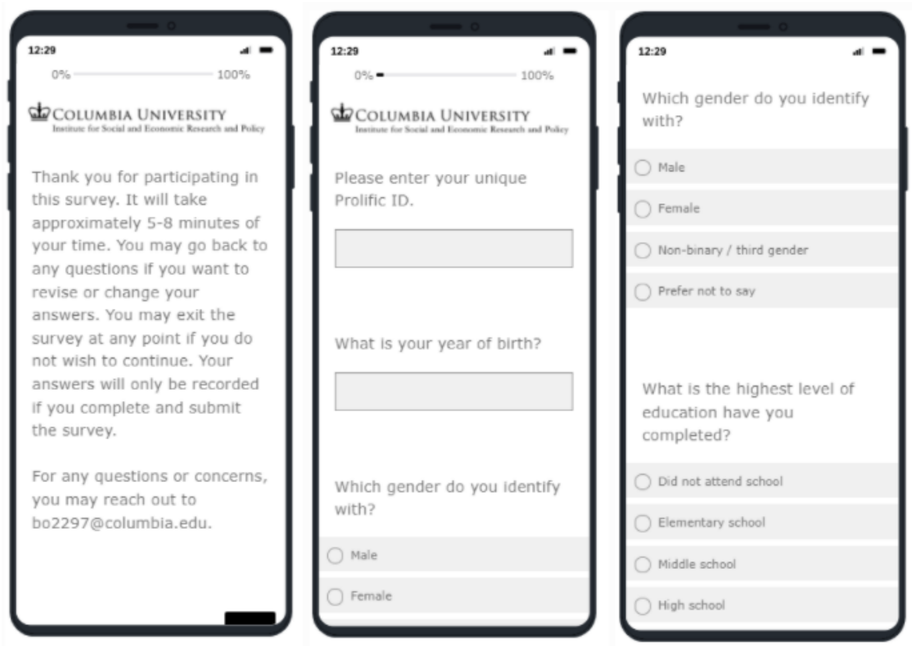


Figure 7: Survey pages 1-3

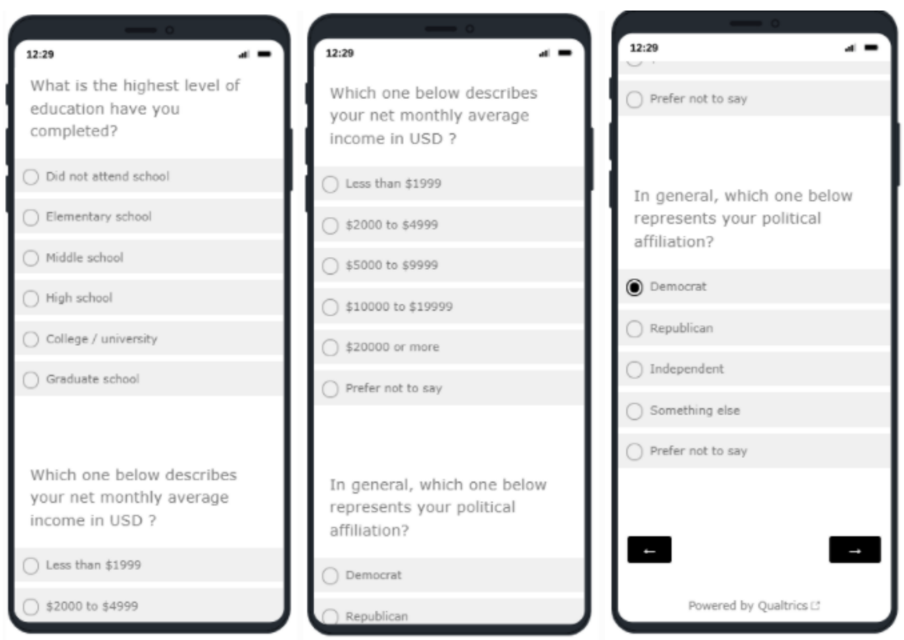


Figure 8: Survey pages 4-6

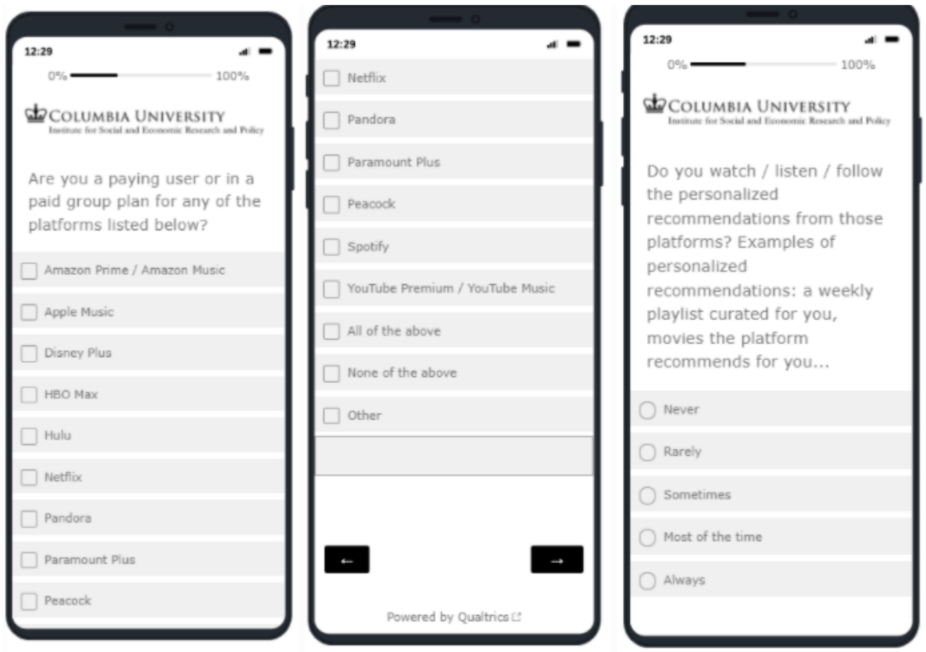


Figure 9: Survey pages 7-9

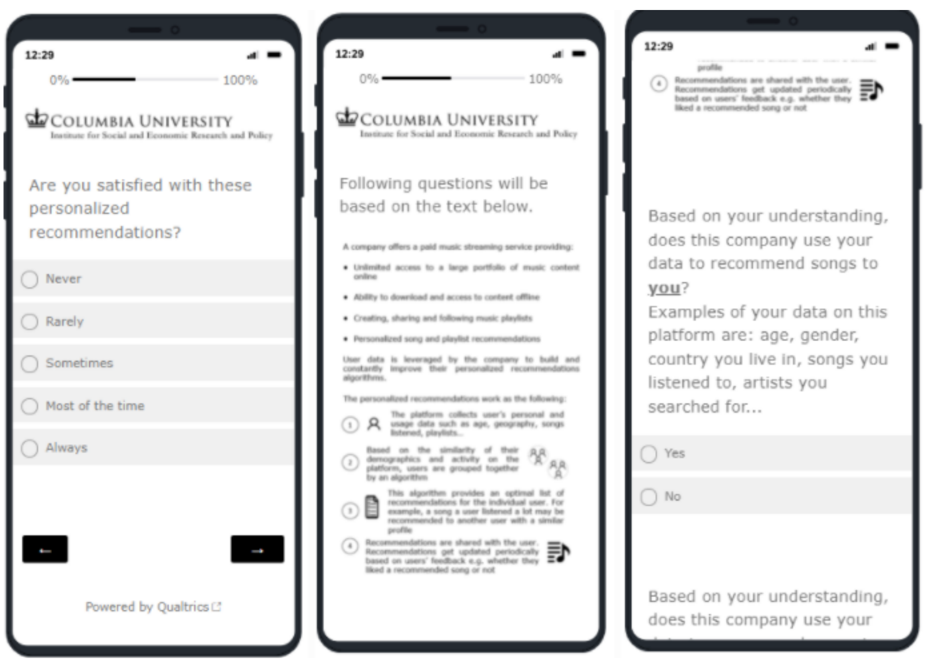


Figure 10: Survey pages 10-12

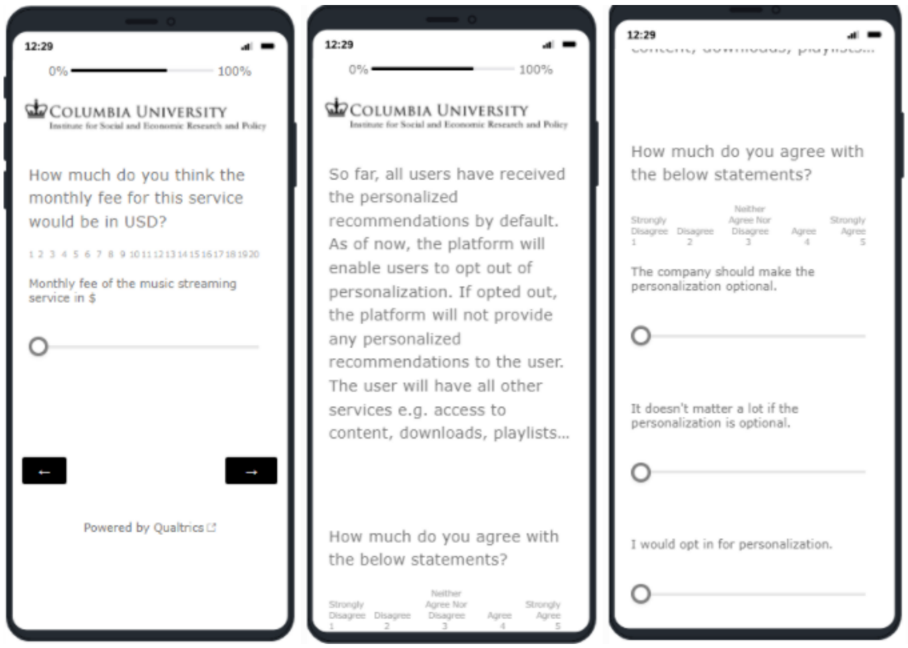


Figure 11: Survey pages 13-15

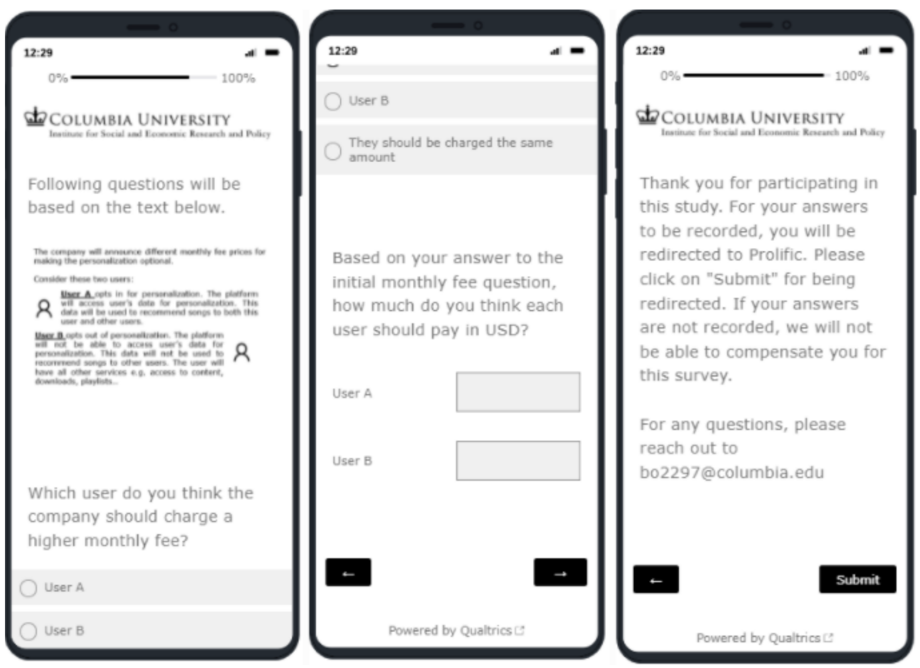


Figure 12: Survey pages 16-18

A company offers a paid music streaming service providing:

- Unlimited access to a large portfolio of music content online
- Ability to download and access to content offline
- Creating, sharing and following music playlists
- Personalized song and playlist recommendations

The personalized recommendations work as the following:

1. The platform collects user's personal and usage data such as age, geography, songs listened, playlists...
2. An algorithm provides an optimal list of recommendations for the individual user
3. Recommendations are shared with the user. Recommendations get updated periodically based on users' feedback e.g. whether they liked a recommended song or not

Figure 13: First information text read by the control group

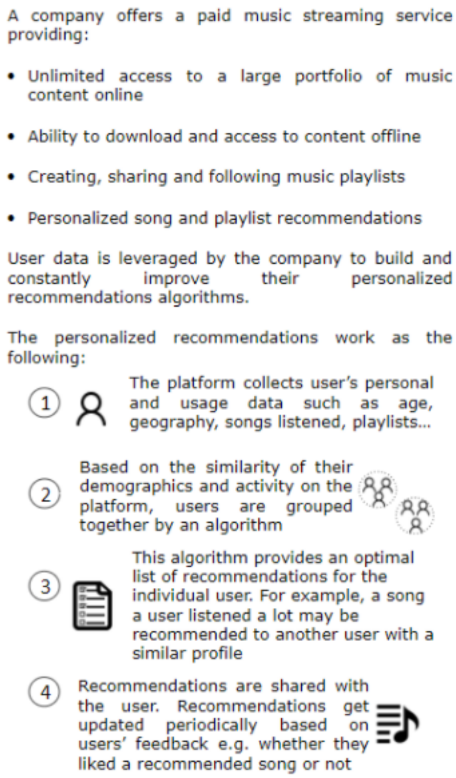
The company will announce different monthly fee prices for making the personalization optional.

Consider these two users:

User A opts in for personalization.

User B opts out of personalization. The user will have all other services e.g. access to content, downloads, playlists...

Figure 14: Second information text read by the control group



A company offers a paid music streaming service providing:

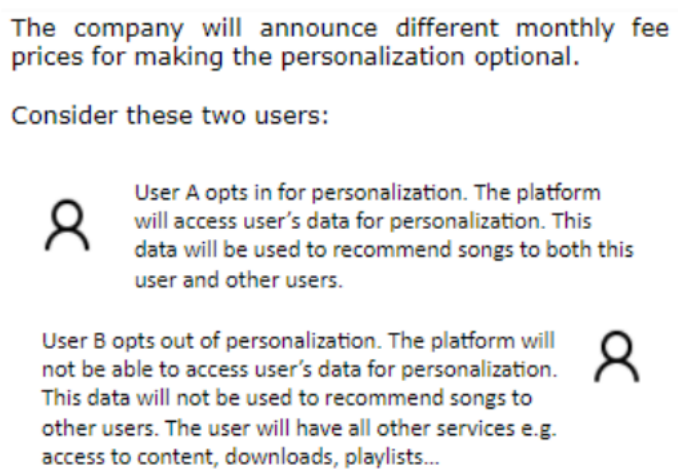
- Unlimited access to a large portfolio of music content online
- Ability to download and access to content offline
- Creating, sharing and following music playlists
- Personalized song and playlist recommendations

User data is leveraged by the company to build and constantly improve their personalized recommendations algorithms.

The personalized recommendations work as the following:

1. The platform collects user's personal and usage data such as age, geography, songs listened, playlists...
2. Based on the similarity of their demographics and activity on the platform, users are grouped together by an algorithm
3. This algorithm provides an optimal list of recommendations for the individual user. For example, a song a user listened a lot may be recommended to another user with a similar profile
4. Recommendations are shared with the user. Recommendations get updated periodically based on users' feedback e.g. whether they liked a recommended song or not

Figure 15: First information text read by the treatment group



The company will announce different monthly fee prices for making the personalization optional.

Consider these two users:

User A opts in for personalization. The platform will access user's data for personalization. This data will be used to recommend songs to both this user and other users.

User B opts out of personalization. The platform will not be able to access user's data for personalization. This data will not be used to recommend songs to other users. The user will have all other services e.g. access to content, downloads, playlists...

Figure 16: Second information text read by the treatment group

12 Glossary

OAI: Online activity information

OPD: Online personal data

PII: Personally identifiable information

PSP: Paid streaming platform

TC: Terms and conditions

Bibliography

Amatriain, Xavier, and Justin Basilico. 2012. “Netflix Recommendations: Beyond the 5 stars (Part 1).” The Netflix Tech Blog. https://netflixtechblog.com/netflix-recommendations-beyond-the-5-stars-part-1-55838468f429.

Angela, Winegar, and Cass Sunstein. 2019. “How Much Is Data Privacy Worth? A Preliminary Investigation.” Journal of Consumer Policy 42, no. 1 (07): 425–440. https://doi.org/10.1007/s10603-019-09419-y.

Arrington, Michael. 2006. “AOL Proudly Releases Massive Amounts of Private Data.” TechCrunch. https://techcrunch.com/2006/08/06/aol-proudly-releases-massive-amounts-of-user-search-data/.

Belleflamme, Paul, and Martin Peitz. 2019. “"Ratings, reviews, recommendations and the consumption of cultural goods.” LIDAM Discussion Papers.

Boyd, Clark. 2019. “How Spotify Recommends Your New Favorite Artist.” Towards Data Science. https://towardsdatascience.com/how-spotify-recommends-your-new-favorite-artist-8c1850512af0.

Brant, Rollin. n.d. “Inference for Means: Comparing Two Independent Samples”.https://www.stat.ubc.ca/~{}rollin/stats/ssize/n2.html.

Brynjolfsson, Erik, Yu "Jeffrey" Hu, and Michael D. Smith. 2006. “From Niches to Riches: Anatomy of the Long Tail.” MIT Sloan Management Review. https://sloanreview.mit.edu/article/from-niches-to-riches-anatomy-of-the-long-tail/.

Budzinski, Oliver. 2003. “Cognitive Rules, Institutions, and Competition.” Constitutional Political Economy 14 (09): 213-233. https://doi.org/10.1023/A:1024754512997.

Budzinski, Oliver, Sophia Gaenssle, and Nadine Lindstädt-Dreusicke. 2021. “Data (R)Evolution – The Economics of Algorithmic Search and Recommender Services.” Ilmenau Economics Discussion Papers, 27, no. 148 (01). http://dx.doi.org/10.2139/ssrn.3769386.

Burgess, Matt. 2021. “All the Ways Spotify Tracks You—and How to Stop It.” Wired. https://www.wired.com/story/spotify-tracking-how-to-stop-it/.

Chhabra, Sameer. 2017. “Netflix says 80 percent of watched content is based on algorithmic recommendations.” Mobile Syrup. https://mobilesyrup.com/2017/08/22/80-percent-netflix-shows-discovered-recommendation/.

Collins, Barry. 2021. “What Your Netflix Data Reveals About You.” Forbes. https://www.forbes.com/sites/barrycollins/2021/08/28/what-your-netflix-data-reveals-about-you.

Confessore, Nicholas. 2018. “Cambridge Analytica and Facebook: The Scandal and the Fallout So Far.” The New York Times. https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html.

De Jong, Menno, and Susanne Barth. 2017. “The privacy paradox – Investigating discrepancies between expressed privacy concerns and actual online behavior – A systematic literature review.” Telematics and Informatics 34, no. 7 (04): 1038-1058. https://doi.org/10.1016/j.tele.2017.04.013.

Dynarisk. n.d. “Netflix, Spotify EA: discover the most ‘cyber-attacked’ brands,” Report by Cybersecurity firm Dynarisk. Dynarisk. https://www.dynarisk.com/resources/blog/discover-the-most-cyber-attacked-brands.

European Union. 2013. “Article 29 Data Protection Working Party.” https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf.

Floridi, Luciano, and Mariarosaria Taddeo. 2015. “The Debate on the Moral Responsibilities of Online Service Providers.” Science and Engineering Ethics 22 (11): 1575-1603. https://doi.org/10.1007/s11948-015-9734-1.

Gran, Anne-Britt, Taina Bucher, and Peter Booth. 2020. “To be or not to be algorithm aware: a question of a new digital divide?” Information, Communication Society 24, no. 12 (03): 1779-1796. 10.1080/1369118X.2020.1736124.

Guynn, Jessica. 2020. “What you need to know before clicking ’I agree’ on that terms of service agreement or privacy policy.” USA Today. https://www.usatoday.com/story/tech/2020/01/28/not-reading-the-small-print-is-privacy-policy-fail/4565274002/.

Heath, Alex. 2015. “How Spotify’s Discover Weekly playlist knows exactly what you want to hear.” Business Insider. https://www.businessinsider.com/how-spotify-discover-weekly-works-2015-9.

Herzog, Stefan M., Anastasia Kozyreva, Philipp Lorenz-Spreen, Ralph Hertwig, and Stephan Lewandowsky. 2021. “Public attitudes towards algorithmic personalization and use of personal data online: evidence from Germany, Great Britain, and the United States.” Humanities and Social Sciences Communications 8, no. N/A (05): Article 117. https://doi.org/10.1057/s41599-021-00787-w.

Hlebowitsh, Nadia. 2021. “50+ Statistics Proving Spotify Growth is Soaring in 2021.” Siteefy. https://siteefy.com/spotify-statistics/.

Hosey, Christine, Fernando Diaz, Zahra Nazari, Jean Garcia-Gathright, and Brian St. Thomas. 2018. “Understanding and Evaluating User Satisfaction with Music Discovery.” The 41st International ACM SIGIR Conference, (07), 55-64. 10.1145/3209978.3210049.

Hunt, Neil, and Carlos A. Gomez-Uribe. 2015. “The Netflix Recommender System: Algorithms, Business Value, and Innovation.” ACM Trans. Manage. Inf. Syst. 6, no. 4 (12): Article 13. http://dx.doi.org/10.1145/2843948.

Jannach, Dietmar, and Michael Jugovac. 2019. “Measuring the Business Value of Recommender Systems.” ACM Trans. Manage. Inf. Syst. 10, no. 4 (12): Article 16

Kokolakis, Sypros. 2017. “Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon.” Computers Security 64 (01): 122-134. https://doi.org/10.1016/j.cose.2015.07.002.

Lakens, Daniel. 2015. “Always use Welch’s t-test instead of Student’s t-test.” The 20% Statistician. http://daniellakens.blogspot.com/2015/01/always-use-welchs-t-test-instead-of.html.

LaMorte, Wayne W. 2016. “Central Limit Theorem.” Boston University School of Public Health Web Resources. https://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704$_{p}robability/BS704_Probability12.html$.

Ljubljana School of Busines. 2020. “Modern Approaches to Knowledge Management Development.” https://dspace.uzhnu.edu.ua/jspui/bitstream/lib/35310/1/Collective%20monograph.Slovenia.%20Publishing%20LJUBLJANA%20SCHOOL%20OF%20BUSINESS%202020.pdf#page=112.

Madani, Ali. 2021. “Machine Learning - Groups are collections of similar individuals.” Cyclicarx. https://blog.cyclicarx.com/machine-learning-groups-are-collections-of-similar-individuals.

Maheshwari, Sapna. 2017. “Netflix and Spotify are discovering that some people find data mining for ads creepy.” Financial Post. https://financialpost.com/technology/netflix-and-spotify-are-discovering-that-some-people-find-data-mining-for-ads-creepy.

Maple, Carsten, Anya Skatova, Rebecca McDonald, and Sinong Ma. 2019. 1Unpacking Privacy: Willingness to pay to protect personal data. 10.31234/osf.io/ahwe4.

Markman, Jon. 2019. “Netflix Harnesses Big Data To Profit From Your Tastes.” Forbes. https://www.forbes.com/sites/jonmarkman/2019/02/25/netflix-harnesses-big-data-to-profit-from-your-tastes.

Mejia, Hiro. 2020. “Market Research: Spotify.” https://uploads-ssl.webflow.com/5e16e5818479800c3ed2aa4c/5ec62a70fe35e282d5b18576_Market%20Research%20-%20Spotify%20(1).pdf.

Netflix. n.d. “How Netflix’s Recommendations System Works.” Netflix Help Center. https://help.netflix.com/en/node/100639/us.

Obar, Jonathan A., and Anne Oeldorf-Hirsch. 2018. “The Biggest Lie on the Internet: Ignoring the Privacy Policies and Terms of Service Policies of Social Networking Services.” Information, Communication Society, (06), 1-20. http://dx.doi.org/10.2139/ssrn.2757465.

Paul, Kari. 2018. “What does your Netflix viewing history reveal about you?” Market Watch. https://www.marketwatch.com/story/heres-what-your-netflix-history-reveals-about-you-2018-09-07.

Prey, Robert. 2018. “Nothing personal: algorithmic individuation on music streaming platforms.” Media, Culture Society 40 (7): 1086–1100. 10.1177/0163443717745147.

Raine, Lee, Brooke Auxier, Monica Anderson, Andrew Perrin, Madhu Kumar, and Erica Turner. 2019. “Americans and Privacy: Concerned, Confused and Feeling Lack of Control Over Their Personal Information.” Pew Research Center. https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information/.

Romm, Tony. 2020. “Amazon, Apple, Facebook and Google grilled on Capitol Hill over their market power.” The Washington Post. https://www.washingtonpost.com/technology/2020/07/29/apple-google-facebook-amazon-congress-hearing/.

Senecal, Sylvain, and Jacques Nantel. 2004. “The Influence of Online Product Recommendations on Consumers’Online Choices.” Journal of Retailing, (08), 159-169. 10.1016/j.jretai.2004.04.001.

Singleton, Jr., Royce A., and Bruce C. Straits. 2017. “Survey Research.” In Approaches to Social Research, 256-298. 6th ed. New York: Oxford University Press. Columbia University Libraries.

Spotify. 2020. “Spotify Users Have Spent Over 2.3 Billion Hours Streaming Discover Weekly Playlists Since 2015.” For the Record. https://newsroom.spotify.com/2020-07-09/spotify-users-have-spent-over-2-3-billion-hours-streaming-discover-weekly-playlists-since-2015/.

Spotify Inc. 2021. “Spotify Privacy Policy.” https://www.spotify.com/us/legal/privacy-policy/.

Stassen, Murray. 2020. “Spotify Invites Record Labels Into Spotify for Artists, Shutters Spotify Analytics.” Music Business Worldwide. https://www.musicbusinessworldwide.com/spotify-invites-record-labels-into-spotify-for-artists-shutters-spotify-analytics/.

STHDA. n.d. “Unpaired Two-Samples T-test in R - Easy Guides - Wiki.” STHDA. Accessed December 18, 2021. http://www.sthda.com/english/wiki/unpaired-two-samples-t-test-in-rinfos.

Tsukayama, Hayley. 2020. "Why Getting Paid for Your Data Is a Bad Deal." Electronic Frontier Foundation. https://www.eff.org/deeplinks/2020/10/why-getting-paid-your-data-bad-deal.

Vanberg, Viktor J. 2002. "Rational Choice vs. Program-based Behavior: Alternative Theoretical Approaches and their Relevance for the Study of Institutions." Rationality and Society 14, no. 1 (02): 7-54. https://doi.org/10.1177/1043463102014001002.

Vigderman, Aliza, and Gabe Turner. 2021. "The Data Big Tech Companies Have On You." Security.org. https://www.security.org/resources/data-tech-companies-have/.

Wallsten, Scott, and Jeffrey Prince. 2020. "How Much is Privacy Worth Around the World and Across Platforms?" Technology Policy Institute. https://techpolicyinstitute.org/wp-content/uploads/2020/02/Prince_Wallsten_How-Much-is-Privacy-Worth-Around-the-World-and-Across-Platforms.pdf.

Wong, Alia. 2018. "What you do on Spotify is public and can be used against you. Here's how to add some privacy." USA Today. https://www.usatoday.com/story/tech/news/2018/07/20/what-you-do-spotify-public-and-can-used-against-you/801647002/.

Wu, Tim. 2015. "Facebook Should Pay All of Us." The New Yorker. https://www.newyorker.com/business/currency/facebook-should-pay-all-of-us.

Zhang, Baobao, and Allan Dafoe. 2019. "Artificial Intelligence: American Attitudes and Trends." Future of Humanity Institute, University of Oxford.

Zylberberg, Hugo, and Nikolas Ott. 2016. "Cyberspace: Explaining How the European Union Is Redefining Ownership and Policies of Personal Data beyond National Borders. Kennedy School Review." Kennedy School Review 16, no. 1 (N/A): 69-75. https://www.proquest.com/docview/2188097924?pq-origsite=gscholar&fromopenview=true.