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

Bengusu Ozcan

Abstract: Paid streaming platforms are not researched as much as online advertising platforms when it comes to online personal data, but utilizing usersâ data is as critical for their personalized services as well. 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 that they already pay for may change, hence these platforms should consider personalization as an optional feature. Our results did not find strong evidence towards that, and indicate that studies about online personal data may not reflect the true behavior of digital servicesâ users in day to day life, which is addressed by acceptability gap or privacy paradox in the literature.

Key Terms: personalization algorithms, data priavacy, personal data, acceptability aap, 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

Literature Review

Online Advertising and Personal Data Exchange

Companies like Facebook, Google and Amazon, so-called âbig techâ, have been in the center of public and legislative debates from online data privacy discussions to antitrust. Starting with the Cambridge Analytica Scandal, in which Facebook allowed a third party to access their usersâ friendsâ 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 can be for any purpose from political propaganda to 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 features and tendencies that this data reveals about them. As Vigderman and Turner provided a recent 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 searched or videos watched (Vigderman and Turner 2021). In this paper, I will refer to all the data a user generates on online platforms as their online personal data (OPD), which covers PII, demographic data and OAI. This data then is leveraged for third parties to give 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 etc.. These companies provide different services from search engines to social media platforms, but as a common business model, they offer their service to users for free in exchange for monetizing their data for online advertising or sponsored content. The algorithms processing the user data for this purpose work behind the scenes and are not visible to users while using the services. (Gran, Bucher, and Booth 2020). In other words, the user generates value for these platforms through this data exchange without being fully aware 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.

Compensation for Sharing or Privacy of Online Personal Data

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. One discussion this exchange brings is whether the companies still owe users a certain share of their profit off of the user data, given that the companiesâ cumulative benefit may be disproportionately higher than the individualsâ benefit off of the platformsâ services, which is a discussion primarily brought into public attention by Tim Wu (Wu 2015). This suggestion may sound fair and incentivize companies to minimize the amount of data they collect only to absolutely necessary ones, however 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) as in the following quote, suggesting that estimating a monetary value for user data undermines individualsâ right to their full data privacy, which EFF defends should be recognized as a fundamental human right (Tsukayama 2020):

> Data dividends will likely be most attractive to those for whom even a small bit of extra money would do a lot. Those vulnerable peopleâlow-income Americans and often communities of colorâshould not be incentivized to pour more data into a system that already exploits them and uses data to discriminate against them.

The other side of the coin is approaching online privacy as an additional service that users may need to pay for in practice, 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 did not change when the benefits they receive from the platforms in exchange of their data were explicitly reminded to respondents (Wallsten and Prince 2020). Another study conducted only in a US sample asked usersâ willingness to pay for privacy of their OPD and also how much they would demand a company to pay them to access their data, across different data types including PII and demographic data that covered their age, sex, household income and political affiliation. Interestingly, the survey found that people expressed a higher amount to pay to protect and being paid to share their demographic data more than PII, which shows that despite demographic data not revealing the identity of a user as precisely as PII can, users still consider their demographic information as sensitive. (Angela and Sunstein 2019). Finally, another study on a smaller sample in the UK about willingness to pay for different OPD types tried to elicit the importance of data by asking users to rank their willingness to share the data and its importance separately, which reveals a strong reverse correlation between importance and willingness to share (Maple et al. 2019). All these studies on willingness to pay for OPD privacy and being paid for the sharing of OPD indicate that all 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 firstly fully informed of the extent to which their online data would be used and secondly, evaluate whether the benefit they get in return of sharing OPD would be worth sharing.

Another criticism that is applicable to both the ideas of platforms paying users for accessing their OPD or users paying for the privacy of their OPD comes from a concept that is referred to as the âacceptability gapâ which simply explains that users miss the connection of providing their OPD and receiving personalized services which is necessary for user to have a benchmark in order to decide whether to give access to their OPD or not. A survey study 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). Herzog et al. further discusses 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 âprivacy paradoxâ, which is explained by Barth and De Jong as (De Jong and Barth 2017):

> â... discrepancies 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.â

The privacy paradox has also been criticized as a concept, because the reported and actual behavior of people online can diverge a lot. It is also tricky to assume that peopleâs OPD privacy attitudes are applicable to their decisions in daily life because they may report their concerns more amplified or directed in an academic survey setting. (Kokolakis 2017).

To sum up, there is no consensus on whether users should receive compensation for sharing their OPD, or be able to pay for an optional increased privacy of their OPD in terms of academic research or industry best practice. Moreover, there are also objections that either of these scenarios would undermine usersâ rights on controlling their OPD. Given the criticisms and also concepts such as acceptability gap that addresses a certain level of lack of user understanding in OPD processing by online platforms, these users may need more transparency and explicit information to understand and evaluate the exchange of their OPD and the equivalent service they would get, and have the option of confirming or rejecting this exchange.

Paid Streaming Platforms and Their Benefit from Online Personal Data

We have discussed OPD processing and findings on usersâ preferences on OPD primarily for the digital platforms that provide a free service in exchange for processing usersâ OPD. 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. 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 usersâ 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 a particularly insightful setting to explore the topics about 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 the awareness of usersâ on their OPD collection by these platforms.

PSPs do not explicitly reveal how their personalization algorithm works given that it is their intellectual property, however there are high level explanations provided by PSPsâ official websites related resources. For example, Netflixâs official customer support website has a section which outlines how the personalized recommendations work, and lists the following as a factor taken into account: âother members with similar tastes and preferences on our serviceâ . 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 PSP users, is a commonly used approach in machine learning (Madani 2021). Therefore, even without an official explanation made by PSPs, it is highly likely that PSP usersâ OAI and demographic information is used to determine what to recommend to similar users.

PSPs generate a significant financial value out of personalization, however it is not as measurable as other income streams that digital platforms have, such as subscription or ad 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 also not available. One approach is referring to major PSPsâ public announcements as their estimated earnings through personalized services. Another approach is to estimate the value that personalization generates alone based on its contribution to user retention and therefore linking its value directly to the revenue that comes from subscription (Jannach and Jugovac 2019). Firstly, personalization offers a value circle which increases userâs 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). Not all users would follow the personalized content on PSPs and the ones that follow would enjoy it on different levels. However, 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 â75% of what people watch is from some sort of recommendationâ (Amatriain and Basilico 2012). Similarly in a 2017 interview, Netflix shared that â80 percent of subscribers trust and follow the recommendations of the algorithmâ (Chhabra 2017). To put it on a financial frame, according to a 2015 study conducted by Netflix, personalization saved them $1B annually by user retention and increasing user engagement on the platform (Hunt and Gomez-Uribe 2015) which may likely be a bigger number given that the platform continued growing its user base during the 7 years passed after this study. The picture is very similar when it comes to online music streaming giant Spotify. The company offers a âDiscover Weeklyâ playlist for its users, which is a weekly updated list of songs that the platformâs recommendation algorithm curates for the user, which has been followed or played by 40 million users in less than a year of its launch back in 2016 (Prey 2018). This statistics is not as strong as Netflixâs calculations on how personalization helps user retention, however in an official press release in 2020, Spotify addressed Discover Weekly as a âflagshipâ 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 many countries, and 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) (Ljubljana School of Business 2020). Similarly, Spotify has been sharing music streaming and behavior insights with music research companies as well as artists on their platform as an additional service that their usersâ 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 attached to a number but are definitely valuable for PSPs.

To sum up, the majority of the research on usersâ preference 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.

Usersâ Benefit from Personalization on Paid Streaming Platforms

After ample elaboration on how personalization is beneficial to PSPs, we will now explore how users benefit from their OPD in PSPs. I will base my arguments mostly on Budzinski et.al.âs comprehensive paper which examines not only PSPs but also e-commerce and search platforms (e.g. Amazon, Google). This paper uses behavioral economics theories and tries to identify how beneficial personalized search and recommendations can be for the users of these platforms (Budzinski, Gaenssle, and LindstÃ¤dt-Dreusicke 2021). The paper starts explaining the most obvious benefits of personalization by referring to other researchersâ findings that personalization saves userâs 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 usersâ buying decisions compared to other online sources of recommendations such as human experts or other consumersâ reviews (Senecal and Nantel 2004).

PSPs do not have as much academic study or market research conducted on their personalization algorithms as much as social media platforms or e-commerce. Most of the available resources about usersâ benefit from personalization on PSPs are limited to small surveys or PSP-led studies. I 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 userâs overall engagement and satisfaction with personalized recommendations (Hosey et al. 2018, p.55-64). A more recent independent research study from

2020 claims that nearly 17% of all playlist listened on Spotify are the ones created by Spotify by personalized recommendations, which is half of the time the general playlists available for all users were listened to as 36% (Mejia 2020). This indicates that majority of the people still prefer listening to playlist curated for a specific theme rather than personalized for them e.g. âhappy road trip songsâ but still a significant amount of time spent on Spotify comes from personalization. Such studies support the theoretical findings that personalization is definitely beneficial for at least a group of users of any online service. In the next section, we will discuss whether there are other types of users who would consider personalization as a compromise on ownership of OPD.

Why Users May Prefer Opting out of Personalization

We mentioned how PSPs publicly announce that their users highly engage with personalized recommendations which they can claim that a considerable amount of PSP users may prefer personalized recommendations 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. refers to the rule-following behavior as a behavioral economics concept, which suggests that consumers may prefer efficiency of personalization for routine activities or decisions while prefer a mix of methods or their own intuition for more outstanding or one-off decisions (Vanberg 2002; Budzinski 2003). From this perspective, we do not know whether consuming content, music or video, is a routine or an outstanding activity for users of PSP. For example, you may use Spotify to save a dinner party at home by playing a curated playlist they made for such an occasion or just listen to anything while working on your paper. On the other hand, some users may be quite picky when it comes to music or movies, they may use PSPs only to access the content in a fair and reliable way, and only consume the content 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 or benefit from personalized recommendations.

The user persona in the previous example who was not using personalization at all may not be the majority of any PSP, therefore, the disparity we claimed may not be a priority for PSPs especially due to the operational difficulty of making personalization optional, which we will mention later on in this section. However, there may be a broad range of users who may still not prefer receiving personalized recommendations based on the net tendency coming from how satisfied they are with personalization versus how important data privacy is for them. In the earlier sections, we introduced multiple studies which found that people were willing to pay for ODP privacy, even if the data was not PII, such as demographic information or fitness tracking data. 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 sensitive information for their users. There are two major arguments which may support usersâ 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 or leakage. PSPs do not share user data with third parties for online marketing purposes unlike online advertising platforms such as Facebook or Amazon. Hence such data privacy concern may arise as a cyber-attack on the PSP or based on the use of the PSP itself. When it comes to the cyber attacks, according to a study by cybersecurity firm Dynarisk, major PSPs Netflix and Spotify are in top cyber attack risk amongst all websites since they hold usersâ paid membership and payment information in their databases (Dynarisk, n.d.). OPD user for personalization is not directly linked to paid membership or payment data on PSPs, however, it does not mean that a cyberattack would not target this type of information or leak this type of data along with targeted data. When it comes to PSPsâ own use, the amount and detail of ODP these platforms can have on users raise concern on the extent of these platforms using this data beyond personalization purposes. As a one-off example, In 2017, Netflix has posted on Twitter the following for marketing purposes âTo the 53 people whoâve watched A Christmas Prince every day for the past 18 days: Who hurt you?â. The content of this tweet did not have any malicious intention, was analyzed solely by Netflix based on the data that it owns and hosts legitimately on its servers 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 track every single action on the platform and microtargeting people (Paul 2018). Similar marketing activities in the form of anonymous posts about usersâ OAI have been used by Spotify as well and claimed by the company that have not received any negative sentiment from their users (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 pioneer in the 1990s had a massive reputation loss in 2006 when they leaked 650,000 usersâ search queries which released very sensitive information about some inappropriate or illegal search terms and partially could be traced back to individuals by combining with search terms containing their PII (Arrington 2006).

The second argument we will cover is the principle of purposeful data collection, as the concept that forms the basis of the European Unionâs General Data Protection Regulation (GDPR), which limits the user data collected by digital platforms strictly to the data that the platform needs for processes explicitly shared with the user (European Union 2013). Considering users who do not prefer having personalized services, default collection of user data utilized in personalization algorithm conflicts with this approach in two ways.

Firstly, the information of how their data is utilized for the personalization algorithms may not be explicitly laid out by the platform. We will mention this soon in the next section where we discuss how effective terms and conditions of digital platforms are for informing users on data processing. Secondly, for users who do not prefer having personalization, utilizing their data for the personalization algorithms is a direct conflict with the purposeful data collection principle.

Given that personalization is the default mode 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 their personalization algorithms. Namely, if you have a great taste in music and know some genres very well, you will make your fellow PSP users who do not have as much knowledge but interest in that genre very happy, because they may be recommended to listen to other songs if they have matching songs with your OAI history. The disparity we would like to emphasize here is, Spotify 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 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.

Usersâ Limited Control of Their Data in Paid Streaming Platforms

A foundational principle in purposeful data collection is 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 as frequently asked questions on their official website. For example, Spotify TC has a separate section on data privacy which mentions what type of ODP is collected from users and for what purpose. One caveat with their approach is, Spotify does not explain that usersâ OPD would be utilized by the company for improving their personalization service in general, but rather frames it as âthe data collected for providing you a personalized experienceâ as if userâs data is used only for their 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 on their official customer support website to explain how a userâs 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 their 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 study 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 or the format that the T Cs are being too long 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 to focus on the consequential lack of user awareness on OPD collection and processing as a direct outcome of default personalization provided by PSPs and lack of an option for opting out of personalization.

Major PSPs currently provide users a certain level of control on the extent of their OPD usage and privacy, but it does not reach a full personalization opt-in or opt-out. Spotify is a platform where users can follow each other and if so, view what each other is listening on the platform 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, this does not prevent Spotify from logging usersâ 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 opting out of 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 currently provides users a full and certain ability to opt out from personalization algorithms and necessary data collection for good or until a certain time that user prefers. Building on top of the fact that not many users carefully read and consider TCs, default and forced modes of personalization may prevent users from being aware of such use and collection 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 PSPsâ responsibility to obtain a data-privacy-first design. As a comprehensive critique of all online service providers (OSPs), Floridi et.al. claim that OSPsâ design forces users to share as much data as possible and therefore it is OSPsâ moral responsibility to safeguard OPD privacy and provide users explicit means of modifying their OPD privacy (Floridi and Taddeo 2015). Floridi et.al.'s review mostly focuses on platforms that leverage online advertising, however, as PSPsâ benefit from their usersâ OPD is already introduced in this article, we can claim that same principle would apply for PSPs. 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 companiesâ 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 usage of OPDs primarily for social media sites, and may even assume that since they pay for the platforms, personalization on PSPs 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 may have a particular responsibility to inform users on how their personalization algorithm works based on usersâ OPD as explicitly as possible.

Operational Difficulties of Optional Personalization

Finally, we should also consider how providing personalization as an optional service (POS) may be challenging for PSPs to implement. Firstly, we should mention that major PSPs as seen on Spotifyâs and Netflixâs official websites, position themselves as personalized experience platforms, which means that they position personalization as their default service. If so, POS could be implemented pre-registration, with an explicit or stand alone disclaimer apart from TCs and ideally easy to understand, such as an illustration that outlines how OPD is utilized for personalization and asks for users to report their preference during subscription. If not designed optimally, such disclaimer may still not be clear enough for users and the interruption in the smooth registration process can cause some users not to sign up for the service or feel dissatisfied with user experience. Alternatively, a post-registration design, registration window can be kept as is with the default service being without personalization, and an explicit disclaimer and set of reminders can follow up with the user to manage their personalization preferences from their user control dashboard. This may minimize the user churn at the registration step, however can also lead to a decrease in users who would not opt in for personalization even if they would prefer trying if they do not follow such reminders or feel skeptical of such suggestion by the company. Overall, we should recognize that POS may cause PSPs to lose users at various steps and decrease the number of people using personalized services, which is a differentiating factor of PSPs and also a way of improving their algorithms. As Boyd et.al. outlines in their report about how ethical decisions are implemented in technology companies, teams involved in such ethical decision making in technology companies are in a tension of prioritizing the ethical critique processes versus keeping up with the fast-paced culture and delivery expectations of technology companies (Metcalf, Boyd, and Moss 2019). This means that a design decision conflicting with business goals may not easily be approved and implemented internally by the companies. This brings the question of whether POS can be mandated through regulation by state actors, which may bring caveats and inefficiencies. According to a critique by The Brookings Institute, implementing such ethical decision changes through regulation firstly may not be fast enough in adapting the changing technology design and secondly would only be applicable in a certain region given that major PSPs operate globally (Roth and Kearns 2020). A final challenge we will mention is about the pricing of PSP services. Given that PSPsâ service with and without personalization would be distinctly different from each other, users may expect a difference in the service fee that they pay. However, based on the acceptability gap concept that we introduced earlier in this section, it is not straightforward to assume whether users would perceive opting out of personalization as a data privacy add-on that they would pay extra for or as an inferior service that they would pay less than the service with personalization. It is also not straightforward to decide whether there should be a price difference in the first place, because the same critique of willingness to pay for privacy applies here as well. If POS introduces a price difference, it would treat ODP as a commodity and undermine usersâ data privacy rights. There may be design decisions that overcome the challenges we mentioned above and even lead to new technology innovations, which will not be detailed for the purposes of this paper. We will focus on what would be the usersâ perceptions towards POS in PSPs in terms of preferences and the applicability of the concept of willingness to pay for data privacy.

Literature Review Summary

So far, we have explained how a user's OPD may be valuable to the user beyond PSPsâ personalization services, and as a popular debate that addresses this value compensation and transparency challenge, we mentioned that users paying for privacy of their OPD or PSPs paying users for accessing their OPD may compromise user awareness on personal data ownership. We then introduced POS as a design that PSPs can adopt to address this challenge, however also briefly mentioned the implementation challenges of such a design change. The purpose of this paper is to introduce the explicit information on how PSPs leverage OPD for personalization as a treatment factor and measure whether users perceive the value they receive from PSPs differently. Would users of PSPs be willing to pay for the PSP more or less than how much they would pay when they did not know how their ODP would contribute to PSPs beyond the fee that they pay? Moreover, if users who are already paying for these services were offered to opt out of personalization, would they perceive it as a data privacy improvement and pay more or a service downgrade and pay less?

If answered, these questions would bring valuable insights for defining into the discussion of the value of personal data, given that paid online streaming platforms bring a different and exceptional setting than the social media platforms. We argue that the value that paid streaming platforms generate out of usersâ data in the means of personalization may not be visible to users, and if informed, their preference on receiving personalized recommendations or what is the fair amount they should be willing for such service may change significantly.

Our Study

This paper builds hypotheses based on the findings about user concerns on OPD and explores these concepts specifically for PSPs given that they are not as widely researched as social media or e-commerce platforms, but essentially different from them by not leveraging online advertising but charging their users directly. Our independent variable is explicit information on how usersâ data is leveraged by PSPs. Our dependent variable is how much the user is willing to pay for such PSP service, representing the perception of value that the user gets from the PSP.

Our first hypothesis 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 company that they make. Therefore, compared to a group of people who do not have the explicit information, the other group with the explicit information may prefer paying a lower fee for the PSP.

Our second hypothesis 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 consider the service with personalization should be more expensive or not from the version without personalization. The additional data privacy that would come from opting out of personalization by a PSP 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 just simply may consider personalization as an additional benefit they are willing to receive. Therefore, we will look for a difference for the difference between price with and without personalization.

Method

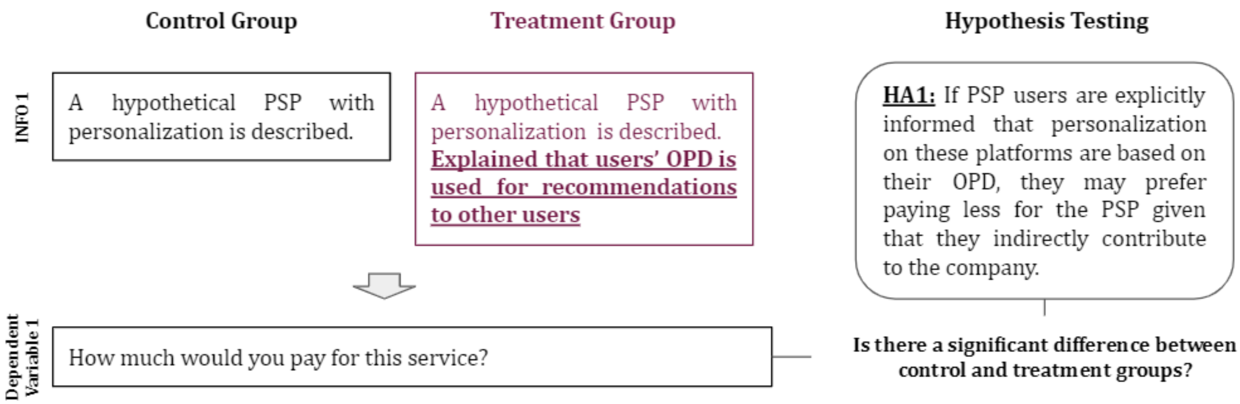
In order to test these hypotheses, a randomized controlled trial (RCT) was designed with one treatment and one control group, all identical except the treatment of providing explicit information on how PSPs leverage OPD for personalization algorithms. For convenience of data collection, the RCT was conducted 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 were introduced to a hypothetical PSP providing music streaming services and personalized recommendations to users. With the intention of avoiding any opinion on existing PSPs priming the responses, a hypothetical platform is preferred rather than referring to an existing platform. 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 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 portion of music content 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 usersâ 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. Questions of the survey and the details of this text along with 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 (D1) 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. The demonstration of this hypothesis testing can be seen in Figure 1.

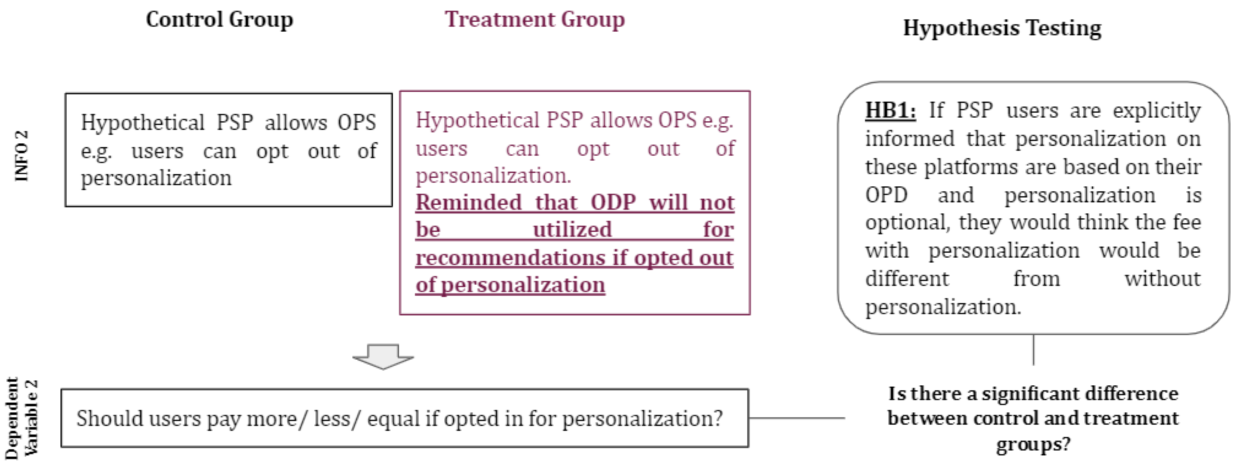
Figure 1: Demonstration of Hypothesis A1



After reading 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, however the treatment group is reminded that if opted out, the PSP would not leverage usersâ ODP any longer, to reinforce the intended treatment effect. After reading the text, the users are asked whether User A and User B should pay the same amount or one should be paying more, and to provide a USD amount for how much each user should pay. The difference between the amount User A and User B should pay was designed to be used as the second dependent variable (D2) 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.

The primary statistical measurement method will be 2-sample t-test to measure the difference of D1 and D2 across treatment and control groups. Moreover, linear and logistic regression will 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.

Figure 2: Demonstration of Hypothesis B1.



Survey Design

The online survey was distributed via Qualtrics. A detailed breakdown of the survey flow and the screenshots of each survey page displayed on an electronic device can be found in Appendix B and C. Survey and question design was based on the principles surfaced in experimental design and survey research sections of Singleton and Straitsâ comprehensive guide on social science research. (Singleton, Jr. and Straits 2017)

The survey started with demographic questions such as highest level of education 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 respondentsâ 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 reads that the company collects their data to recommend them songs, without detailing how the personalization algorithms work and their data is leveraged. Unlike the control group, the treatment group reads that the company leverages usersâ data to find similar users and recommend them the songs that similar users listened to and therefore constantly improves their recommendation algorithm with the help of user data. The exact texts are available in Appendix C.

After the text, there were two understanding checks that asked whether the company leveraged a userâs data to recommend songs to that user and other users respectively. It was expected for the treatment groups to answer both questions with yes. Control groupâs 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 streaming platforms group users for song recommendations.

After the understanding checks, the participants responded to three Likert scale questions which measured how much respondents think that a user themselves, other users and the company benefits from a userâs data on this platform. These questions intended to measure whether the treatment created a difference between control and treatment groupsâ perception on how PSPs benefit from usersâ ODP. Outcomes of these questions are going to be analyzed in Further Analysis under the Results section.

After these questions, the respondents were asked their willingness to pay for this hypothetical PSP in USD, which created D1, the first main dependent variable of the study. A range was provided from $1 to $20 with $1 increment, taking Spotifyâs 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 information text. Both groups read that the hypothetical PSP was now offering 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 was reminded that the PSP would be able to leverage only User Aâs 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 whether the treatment created a difference between control and treatment groupsâ personalization preferences 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 D2, 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 section.

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 Spotifyâs 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 the 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 2 and Appendix 3. 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 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 2 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.

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 usersâ self-identification. According to the self-report gender question in the survey, control groupâs 49% was male, 46% was female, 4% was non-binary and 1% preferred not to say. Treatment groupâs 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 participantsâ 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.

516 participants who used at least one paid streaming platform were asked how much they follow

Figure 3: Participants using only one paid streaming platform

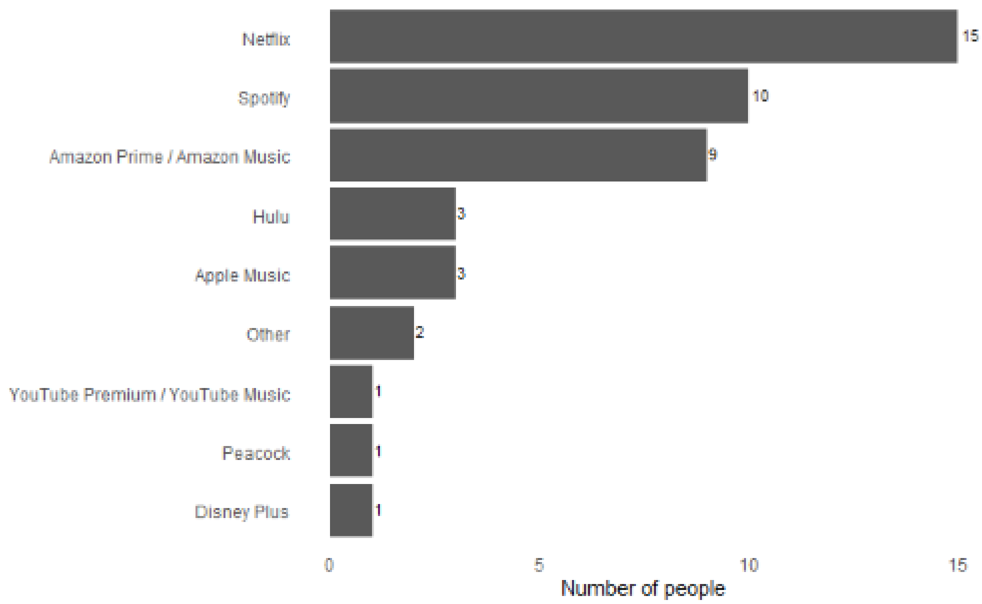
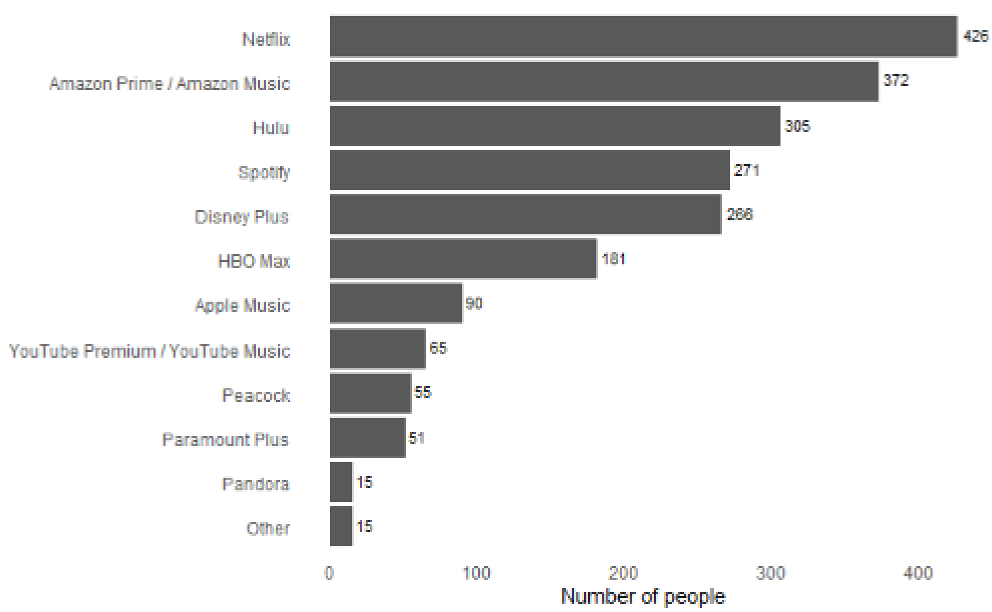


Figure 4: Participants using at least one or more paid streaming platform



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

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%

recommendations. These results can also be seen in Table 2.

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

Results

All results will refer to the short names of the variables introduced in Appendix 1 e.g. Dependent Variable 1 will be called D1. 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.

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, D1. 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 D1 using the Shapiro-Wilk Test and found that D1 is not normally distributed (W=0.93427, p-value $\leq$ 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 D1 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 userâs 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 D2, which was the difference of preferred price of the hypothetical PSP with personalization and without personalization. D2 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 Welchâs t-test for D2, since the latter does not have the equal variances assumption (Lakens 2015).

Both tests resulted very similarly. According to Welchâs 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 D2, 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.

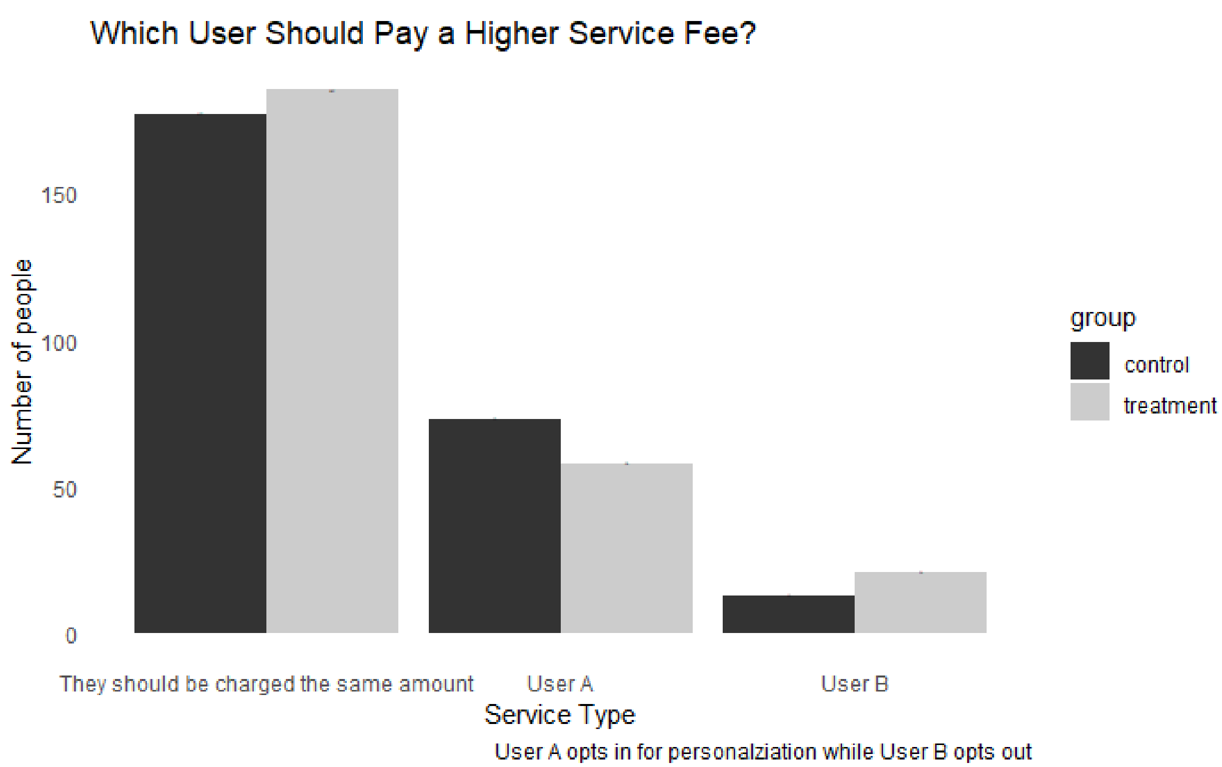
Given that D2 was derived from the difference of D3 and D4, how much users would pay the PSP

Table 4: Responses to D3, D4 and deducting D2 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

with personalization versus without personalization respectively, we looked further into respondentsâ 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.

Figure 5



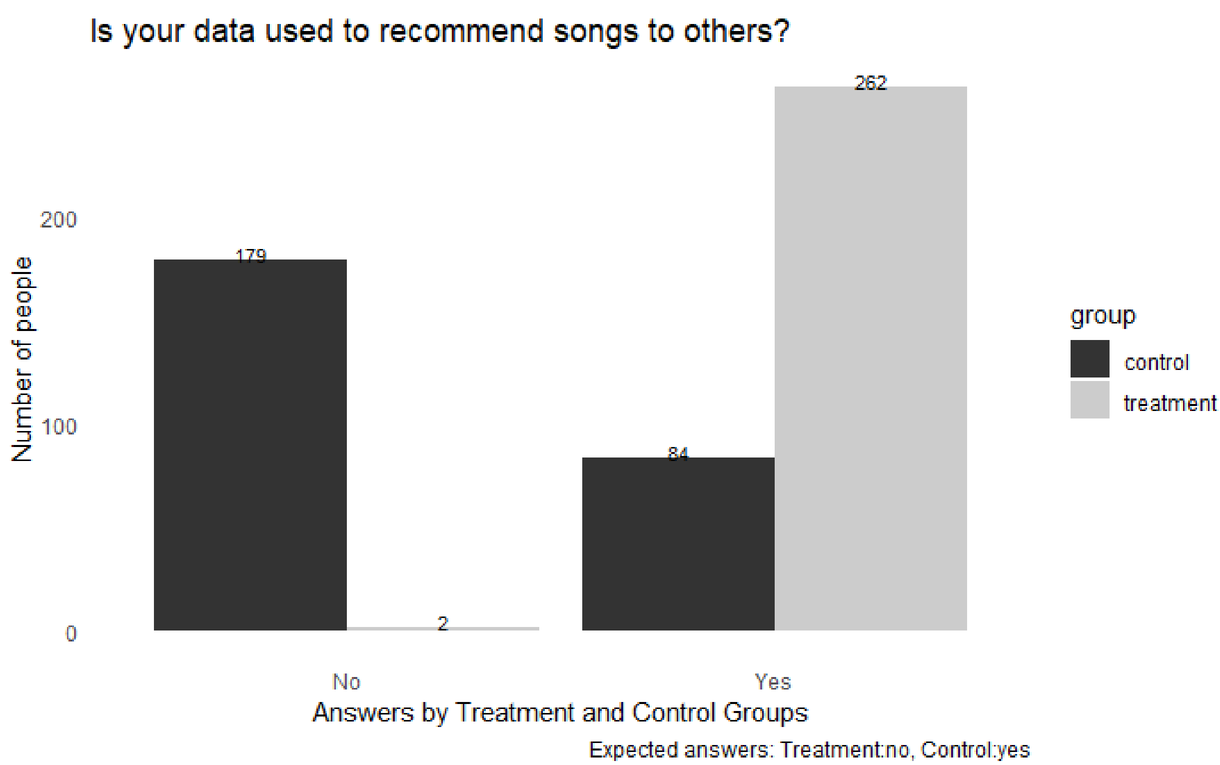
As a supplementary analysis, we applied Welch's t-test for both D3 (t(516.13)=0.53,p=0.6) and D4 (t(521.88)=-0.45,p=0.6),and found no statistically significant difference. Normality and variance tests for these variables are available in Appendix D.

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 usersâ 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 âwhether your data is used for recommending songs to othersâ which treatment group was expected to answer with âyesâ and control group was expected to answer with ânoâ. Figure 6 shows that the majority (68%) of the control group responded ânoâ to this question as expected, however, 32% of them still responded âyesâ 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 âyesâ to the intended treatment effect while 181 of them responded ânoâ, 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.

Figure 6



Applying same variance tests and assuming sufficiently enough large sample, 2-sample t-test did not find any difference on D1 for reconstructed groups. For D2, 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, D2 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 D2 is constructed from D3 and D4, 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 D4(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 D3v(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: D2, D3 and D4 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

Further Analysis

We analyzed both the main dependent variables and the further data on user preferences against the control variables. As an exploratory approach and assuming that n=527 is a big enough sample size, 10 linear regression models are built for all numeric dependent variables listed in Appendix A. Despite the large sample size, since the linear regression has the assumption of normality of the data, logistic regression models are also run as a validation of significance. For this purpose, we created the binary versions of each numeric dependent variable, assigning 0 if the value was smaller than the average of the dependent variable, 1 if equal to or greater than the average. After running both models for the same input and output sets, we introduced only the results that were consistent in terms of statistical significance. Selected results are displayed in Table 6. We interpreted only the variables that were significant in both models. All models are included in the code shared in Appendix D.

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

Regression models supported 2-sample t-tests by showing that being in treatment or control group did not impact D1 significantly. However, both models pointed out that D6 was statistically significant on D1. D6 was how much the respondents agreed with the statement of âOther users benefit from my data on this platform a lot.â on a Likert scale from 1 to 5. For each unit of more agreement, respondents were willing to pay $0.35 less on average for the monthly fee of the service, regardless of all other variables and the treatment factor. This 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 suggests that these variables can explain only 7% of the variance of D1, we still can not claim that the change in D1 can be strongly explained by the treatment factor in the study.

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

The next model is on D5, which is how much the respondents agree with the statement of âAs a user, I benefit from my data on this platform a lot.â from 1 to 5. Both models suggest that the treatment group agreed with D5 significantly less than the control group. This may indicate that when users are aware of how their data benefits other users and PSPs, they may have the perception that they do not receive a sufficiently enough benefit back from the PSP.

Two other findings of this model are linked to each other. More someone stated that they were satisfied with personal recommendations and the more someone agreed with the statement of âI would opt in for personalization.â, respectively for the variables C14 and D10, the more they agreed with the statement in D5. This finding is expected and supports the literature review that a group of PSP users would be very satisfied with personalization services and not as concerned about how their OPD is used as some other users may be.

The significance of D6 on D5 is an expected result of the survey design, because both the treatment and the control groups reported that they believe that the platform leverages their OPD for other users. Moreover, the model suggests that males agree with D5 less than females, which was not expected given that there was no assumption on the gender difference for the results.

The multiple R-square of this model suggests that these variables explain 39% of the variance in D3, which indicates that it is a considerably strong model.

Model3: Optional personalization is preferred if company benefits from users

The next model is on D8, which is how much the respondents agree with the statement of âThe company should make the personalization optional.â from 1 to 5. D5 is found to be statistically significant, which means that the more the respondents agree with the statement of âthe company benefits from my data a lotâ, the more they think that the personalization should be optional. This indirectly supports our assumption in the literature review that a user group which does not benefit from personalization as much as other users may prefer opting out of personalization.

Other two significant variables are D9 and D10, which stand for how much the respondents agreed with the following statements respectively: âIt doesn't matter a lot if the personalization is optional.â, âI would opt in for personalization.â. As expected, these two are inversely correlated with D8. This model has a multiple R-squared of 26%, which makes it considerably strong in explaining the variance of D8.

Model 4: Users satisfied with personalization do not care about optional personalization

The last model is on D10, which is how much the respondents agree with the statement of âI would opt in for personalization.â from 1 to 5. As expected, C14 is statistically significant on D10, which suggests that more satisfied someone with personalized recommendations, more willing they would be to opt in for personalization. Another significant variable D8 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 27%.

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, respondentâs familiarity with PSPs or specifically Spotify were not found to be statistically significant factors

Table 6: Regression Models on D1, D6, D8 and D10

	Dependent variable:							
	D1	Binary D1	D5	Binary D5	D8	Binary D8	D10	Binary D10
	OLS	*logistic*	*OLS*	*logistic*	*OLS*	*logistic*	*OLS*	*logistic*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Group: treatment	0.309	0.354	−0.234***	−0.947***	−0.012	−0.026	−0.080	−0.432*
	(0.313)	(0.218)	(0.075)	(0.265)	(0.076)	(0.241)	(0.112)	(0.231)
Age	0.036	0.008	−0.002	0.007	0.011**	0.036**	−0.011	−0.034**
	(0.023)	(0.016)	(0.006)	(0.019)	(0.006)	(0.018)	(0.008)	(0.017)
Gender:Male	−0.462	−0.021	−0.187***	−0.628***	0.041	0.355	−0.067	−0.238
	(0.285)	(0.198)	(0.069)	(0.234)	(0.069)	(0.218)	(0.102)	(0.209)
Gender:Non-binary / third gender	0.024	0.021	−0.333*	−1.122*	0.275	1.208*	−0.510*	−1.236**
	(0.749)	(0.523)	(0.181)	(0.643)	(0.181)	(0.706)	(0.268)	(0.588)
Gender:Prefer not to say	−3.071	−15.396	−0.248	0.114	0.431	0.994	−0.168	−0.484
	(2.169)	(623.609)	(0.526)	(1.710)	(0.525)	(1.890)	(0.778)	(1.500)
Education_numeric	−0.150	−0.171	0.031	−0.069	−0.078	−0.156	−0.033	0.042
	(0.224)	(0.156)	(0.054)	(0.186)	(0.054)	(0.175)	(0.080)	(0.167)
Income: Less than $1999	−1.214*	−0.687	−0.173	−0.225	−0.147	0.210	0.111	0.124
	(0.631)	(0.444)	(0.153)	(0.512)	(0.153)	(0.487)	(0.226)	(0.464)
Income: $2000 to $4999	−0.509	−0.165	−0.074	0.080	−0.200	−0.078	0.038	−0.013
	(0.602)	(0.425)	(0.146)	(0.487)	(0.146)	(0.461)	(0.216)	(0.439)
Income: $5000 to $9999	−1.351**	−0.667	−0.116	0.118	−0.165	−0.227	0.030	−0.082
	(0.644)	(0.452)	(0.156)	(0.517)	(0.156)	(0.493)	(0.231)	(0.467)
Income: $20000 or more	0.320	0.111	−0.083	0.377	0.005	0.305	0.070	0.226
	(0.805)	(0.577)	(0.195)	(0.731)	(0.195)	(0.611)	(0.289)	(0.601)
Income: Prefer not to say	−0.388	0.027	0.071	0.859	−0.201	−0.546	−0.289	−0.239
	(0.857)	(0.614)	(0.208)	(0.726)	(0.208)	(0.694)	(0.307)	(0.631)
Using_PSPs	−0.319	−0.055	0.028	−0.065	0.062	−0.099	0.010	0.137
	(0.366)	(0.254)	(0.089)	(0.318)	(0.089)	(0.283)	(0.131)	(0.291)
Using_Spotify	0.550*	0.188	0.033	0.231	0.045	0.101	0.038	−0.030
	(0.285)	(0.199)	(0.069)	(0.232)	(0.069)	(0.222)	(0.102)	(0.212)
Personalization_satisfaction	−0.044	−0.006	0.256***	0.634***	−0.013	−0.118	0.160***	0.512***
	(0.142)	(0.099)	(0.032)	(0.120)	(0.034)	(0.103)	(0.050)	(0.103)
Political:Independent	0.280	−0.055	−0.044	0.075	−0.084	−0.448*	0.008	0.084
	(0.310)	(0.215)	(0.075)	(0.254)	(0.075)	(0.244)	(0.111)	(0.231)
Political:Prefer not to say	−1.447	−0.223	−0.629**	−0.674	0.169	−0.382	0.069	−1.716
	(1.270)	(0.882)	(0.307)	(1.124)	(0.308)	(0.965)	(0.456)	(1.197)
Political:Republican	0.083	−0.123	−0.040	0.131	−0.059	−0.392	0.005	−0.258
	(0.413)	(0.284)	(0.100)	(0.346)	(0.100)	(0.326)	(0.148)	(0.305)
Political:Something else	1.184*	0.600	−0.120	−0.533	−0.024	0.210	0.059	0.356
	(0.646)	(0.486)	(0.157)	(0.554)	(0.157)	(0.489)	(0.232)	(0.503)
benefit_1	0.174	0.109			0.018		0.456***	
	(0.184)	(0.128)			(0.045)		(0.063)	
benefit_4	−0.455***	−0.309***	0.201***	0.505***	−0.056	−0.126	−0.037	0.081
	(0.155)	(0.110)	(0.036)	(0.127)	(0.037)	(0.118)	(0.056)	(0.113)
benefit_5	0.283*	0.196*	0.014	0.263*	0.149***	0.483***	−0.006	0.070
	(0.166)	(0.116)	(0.040)	(0.138)	(0.040)	(0.130)	(0.060)	(0.128)
preferences_1	−0.100	−0.083	0.018	0.124			−0.212***	−0.424***
	(0.184)	(0.129)	(0.045)	(0.150)			(0.065)	(0.146)
preferences_4	0.110	0.102	0.010	−0.070	−0.283***	−0.814***	0.111**	0.155
	(0.136)	(0.095)	(0.033)	(0.110)	(0.030)	(0.112)	(0.048)	(0.103)
preferences_5	−0.040	−0.005	0.208***	0.601***	−0.096***	−0.237***		
	(0.124)	(0.086)	(0.029)	(0.099)	(0.030)	(0.090)		
Constant	8.933***	0.565	1.541***	−5.680***	4.723***	0.752	2.553***	0.683
	(1.719)	(1.196)	(0.411)	(1.470)	(0.359)	(1.153)	(0.606)	(1.270)
Observations	527	527	527	527	527	527	527	527
R^2	0.074		0.399		0.261		0.286	
Adjusted R^2	0.030		0.371		0.228		0.253	
Akaike Inf. Crit.		747.614		586.050		633.566		677.365

Note: *p<0.1; **p<0.05; ***p<0.01

in any of our analyses, but the treatment effect may have been different if the survey introduced a different PSP.

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 peopleâs 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 usersâ 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 ODP 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 ODP 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.

Conclusion

Online personal data becomes more of an asset for many digital businesses and there are conflicting findings about peopleâs perception of their data online. Referred to as âacceptability gapâ 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 peopleâs 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 usersâ 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.

Appendices

Appendix 1

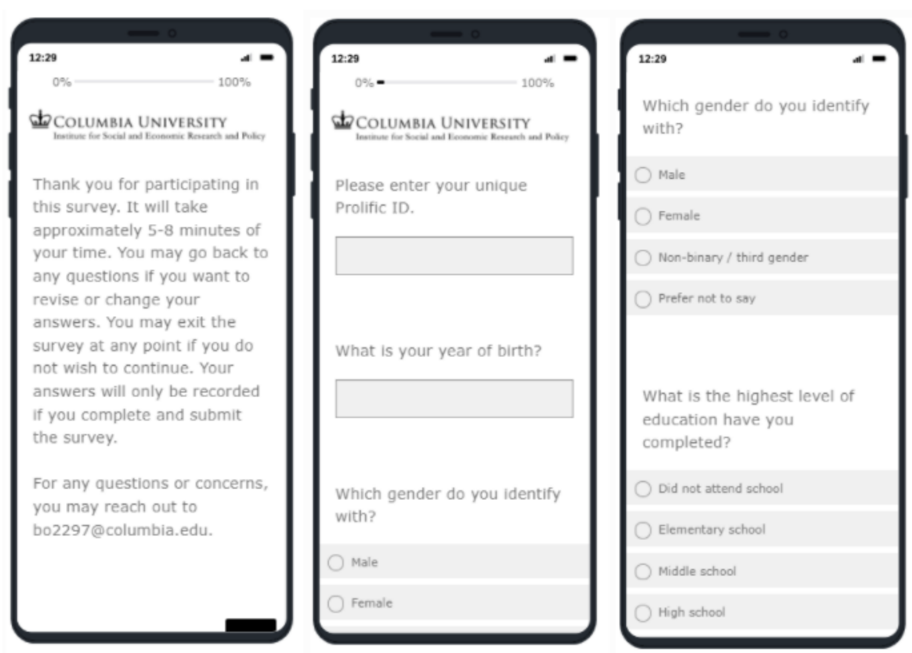
Variable Name	Description	Variable Type	Short Name	Raw Data / Deducted
Age	Age in 2021	Control	C1	Deducted
Gender	Reported gender identification	Control	C2	Raw
Education	Highest level of education completed	Control	C3	Raw
Education_numeric	Highest level of education recoded from 4 to 1, highest to lowest 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	D1	Raw
Personalization_fee_difference	Subtraction of fee_2_2 from fee_2_1	Dependent	D2	Raw
Fee_personalization	Fee with personalization	Dependent	D3	Raw
Fee_without_personalization	Fee without personalization	Dependent	D4	Raw
benefit_1	1 to 5 likert scale on agreement with the following: As a user, I benefit fromÂ my data on thisÂ platform a lot.	Dependent	D5	Raw
benefit_4	1 to 5 likert scale on agreement with the following: Other users benefitÂ from my data on this platform a lot.Â	Dependent	D6	Raw

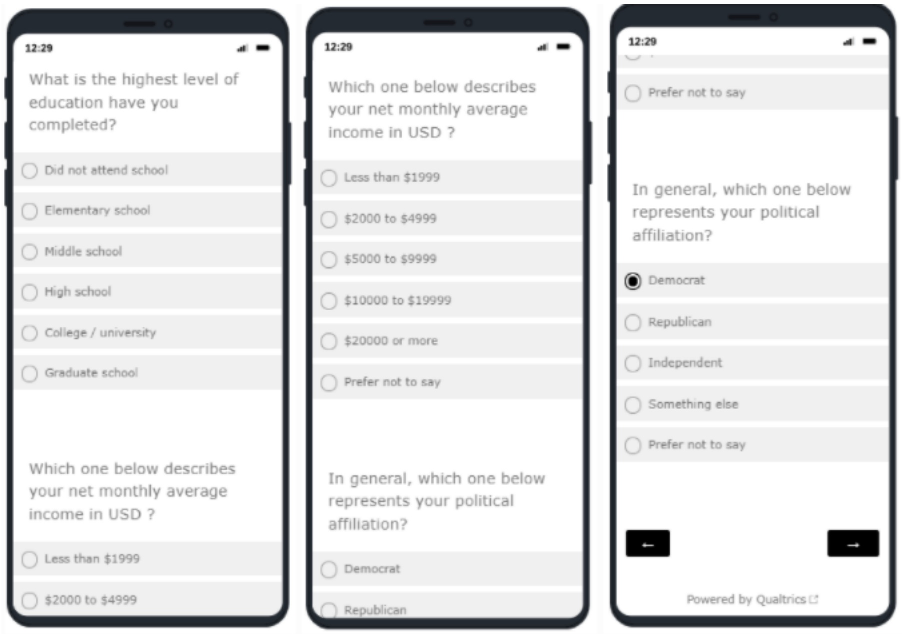
benefit_5	1 to 5 likert scale on agreement with the following: The company benefitsÂ from my data on thisÂ platform a lot.	Dependent	D7	Raw
preferences_1	1 to 5 likert scale on agreement with the following: The company should make the personalization optional.	Dependent	D8	Raw
preferences_4	1 to 5 likert scale on agreement with the following: It doesn't matter a lot if the personalization is optional	Dependent	D9	Raw
preferences_5	1 to 5 likert scale on agreement with the following: I would opt in for personalization.	Dependent	D10	Raw
equal_fee	Whether respondents think the fee with or without personalization should be the same	Dependent	D11	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

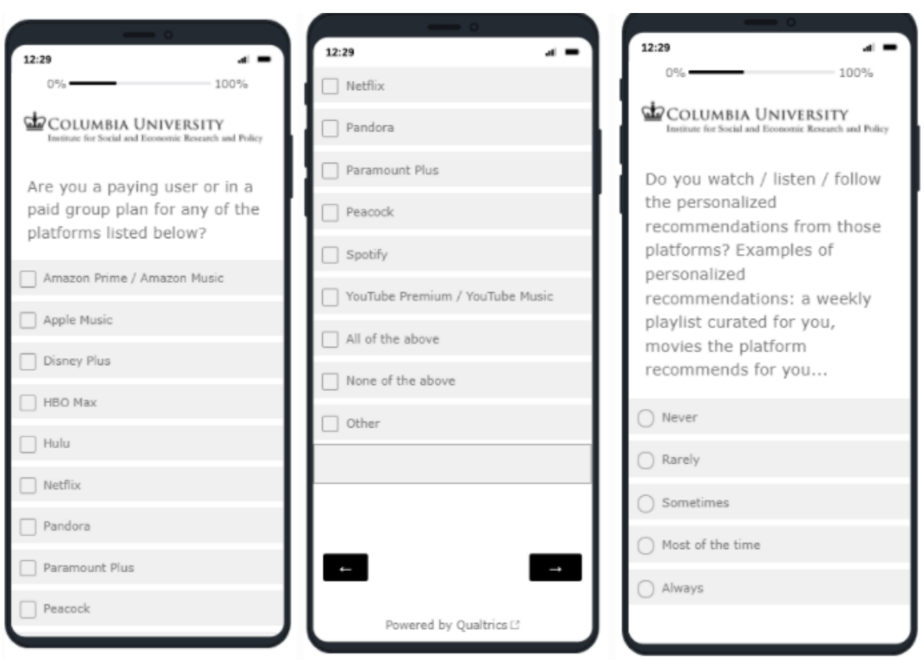
Appendix 2

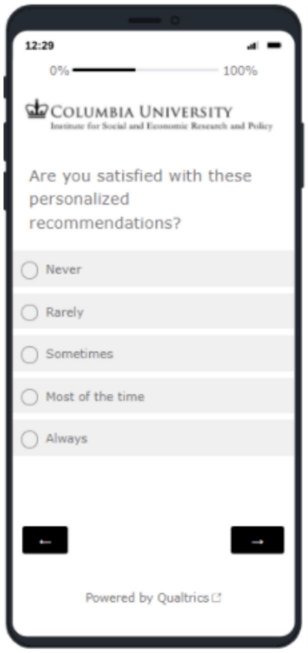
Data files are shared in the public Github repository, Link to detailed survey questions, survey flow and contact details for detailed questions are avaialble in the Readme of the Github repository.

Appendix 3

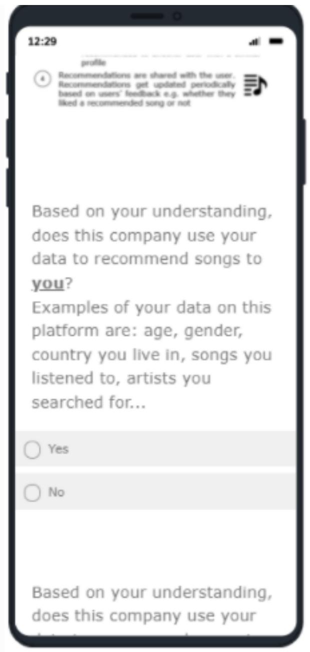


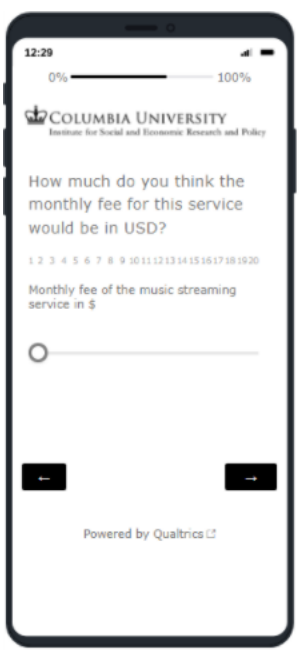

12:29
What is the highest level of education have you completed?
Did not attend school
Elementary school
Middle school
High school
College / university
Graduate school
Which one below describes your net monthly average income in USD ?
Less than $1999
$2000 to $4999
12:29
Which one below describes your net monthly average income in USD ?
Less than $1999
$2000 to $4999
$5000 to $9999
$10000 to $19999
$20000 or more
Prefer not to say
In general, which one below represents your political affiliation?
Democrat
Republican
12:29
Prefer not to say
In general, which one below represents your political affiliation?
Democrat
Republican
Independent
Something else
Prefer not to say
←
→
Powered by Qualtrics

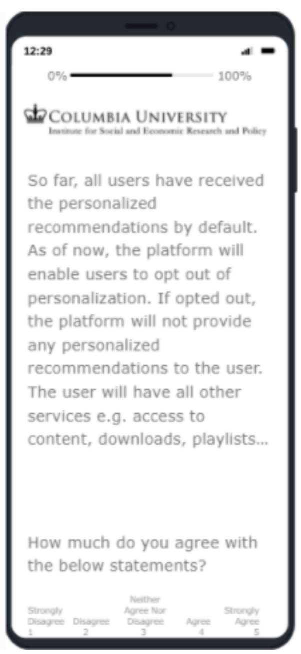

12:29
0%
100%
Columbia University
Institute for Social and Economic Research and Policy
Are you a paying user or in a paid group plan for any of the platforms listed below?
Amazon Prime / Amazon Music
Apple Music
Disney Plus
HBO Max
Hulu
Netflix
Pandora
Paramount Plus
Peacock
12:29
Netflix
Pandora
Paramount Plus
Peacock
Spotify
YouTube Premium / YouTube Music
All of the above
None of the above
Other
←
→
Powered by Qualtrics
12:29
0%
100%
Columbia University
Institute for Social and Economic Research and Policy
Do you watch / listen / follow the personalized recommendations from those platforms? Examples of personalized recommendations: a weekly playlist curated for you, movies the platform recommends for you...
Never
Rarely
Sometimes
Most of the time
Always

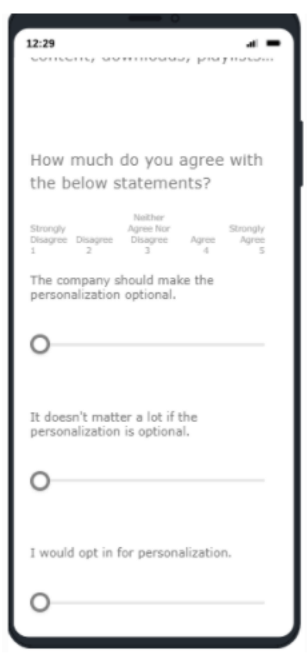

12:29
0% 100%
Columbia University
Institute for Social and Economic Research and Policy
Are you satisfied with these personalized recommendations?
Never
Rarely
Sometimes
Most of the time
Always
Powered by Qualtrics


12:29
0% 100%
Columbia University
Institute for Social and Economic Research and Policy
Following questions will be based on the text below.


12:29
Based on your understanding, does this company use your data to recommend songs to you?
Examples of your data on this platform are: age, gender, country you live in, songs you listened to, artists you searched for...
Yes
No
Based on your understanding, does this company use your


12:29
0% 100%
Columbia University
Institute for Social and Economic Research and Policy
How much do you think the monthly fee for this service would be in USD?
Monthly fee of the music streaming service in $
Powered by Qualtrics


12:29
0% 100%
Columbia University
Institute for Social and Economic Research and Policy
So far, all users have received the personalized recommendations by default. As of now, the platform will enable users to opt out of personalization. If opted out, the platform will not provide any personalized recommendations to the user. The user will have all other services e.g. access to content, downloads, playlists...
How much do you agree with the below statements?


12:29
How much do you agree with the below statements?
The company should make the personalization optional.
It doesn't matter a lot if the personalization is optional.
I would opt in for personalization.

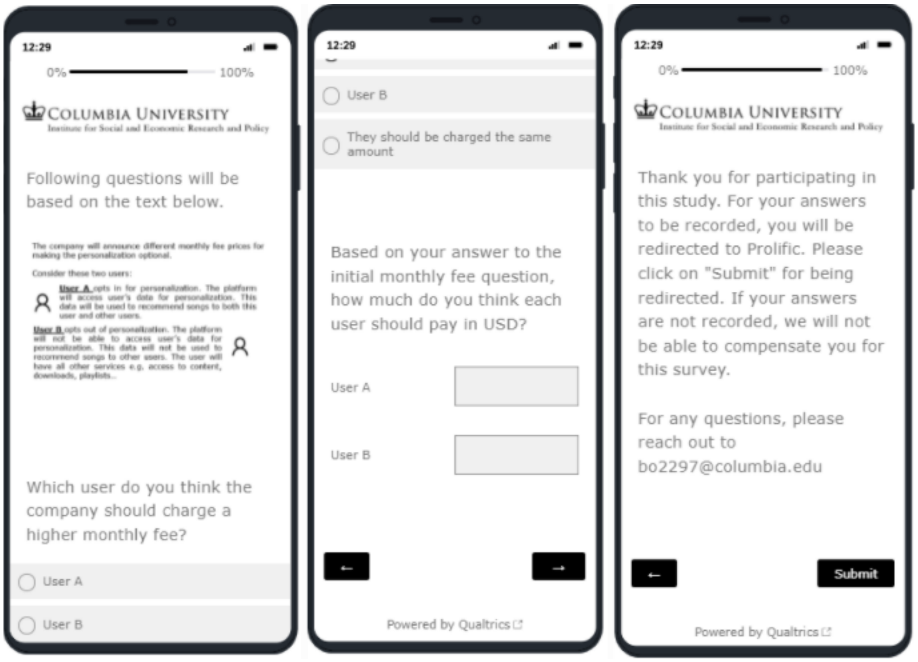
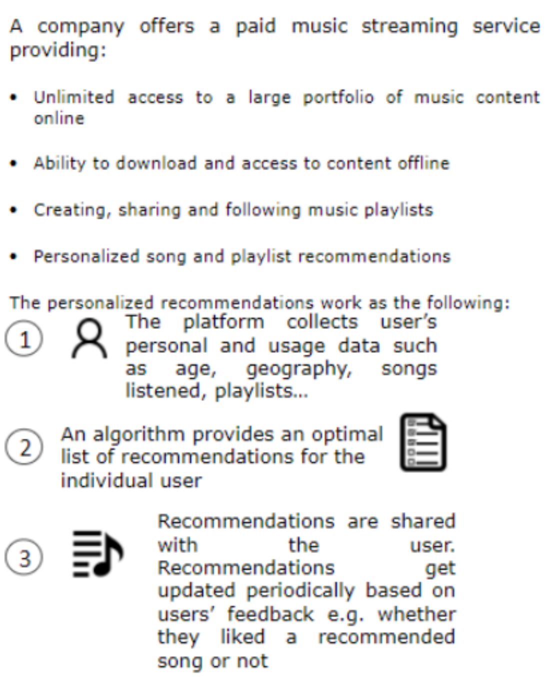


Figure 7: First information text read by the control group:


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

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 8: Second 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 9: First information text read by the treatment group:

A company offers a paid music streaming service providing:

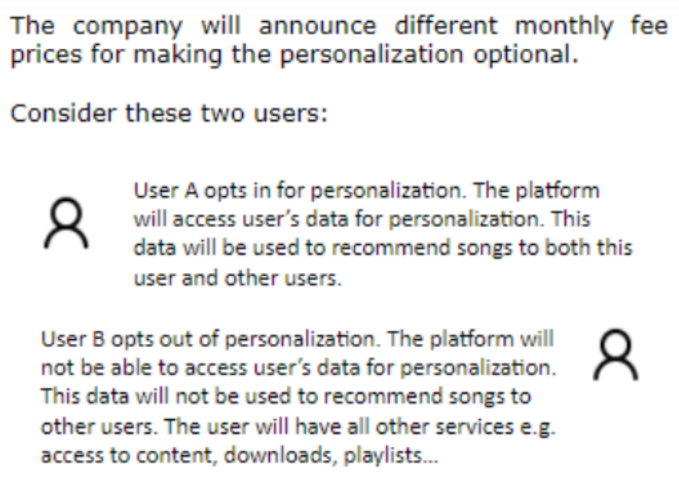
- Unlimited access to a large portfolio of music content online
- Ability to download and access to content offline
- Creating, sharing and following music playlists
- Personalized song and playlist recommendations

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 10: Second 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...

References

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/?sh=739555625f49.

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_probability/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/?sh=259a060766fd.

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

Metcalf, Jacob, Danah Boyd, and Emanuel Moss. 2019. "Owning Ethics: Corporate Logics, Silicon Valley, and the Institutionalization of Ethics." *Social Research: An International Quarterly* 86 (2): 449-476. muse.jhu.edu/article/732185.

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/.

Roth, Aaron, and Michael Kearns. 2020. “Ethical algorithm design should guide technology regulation.” The Brookings Institute. https://www.brookings.edu/research/ethical-algorithm-design-should-guide-technology-regulation/.

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-r#infos.

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. http://ezproxy.cul.columbia.edu/login?url=https://www.proquest.com/scholarly-journals/european-perspective-on-protection-personal-data/docview/2188097924/se-2?accountid=10226.