OtherTube: Facilitating Content Discovery and Reflection by Exchanging YouTube Recommendations with Strangers

Md Momen Bhuiyan
Virginia Tech
USA
momen@vt.edu

Carlos Augusto
Bautista Isaza
Virginia Tech
USA
carlosaugusto@vt.edu

Tanushree Mitra
University of Washington
USA
tmitra@uw.edu

Sang Won Lee
Virginia Tech
USA
sangwonlee@vt.edu

ABSTRACT

To promote engagement, recommendation algorithms on platforms like YouTube increasingly personalize users' feeds, limiting users' exposure to diverse content and depriving them of opportunities to reflect on their interests compared to others'. In this work, we investigate how exchanging recommendations with strangers can help users discover new content and reflect. We tested this idea by developing OtherTube—a browser extension for YouTube that displays strangers' personalized YouTube recommendations. OtherTube allows users to (i) create an anonymized profile for social comparison, (ii) share their recommended videos with others, and (iii) browse strangers' YouTube recommendations. We conducted a 10-day-long user study ($n = 41$) followed by a post-study interview ($n = 11$). Our results reveal that users discovered and developed new interests from seeing OtherTube recommendations. We identified user and content characteristics that affect interaction and engagement with exchanged recommendations; for example, younger users interacted more with OtherTube, while the perceived irrelevance of some content discouraged users from watching certain videos. Users reflected on their interests as well as others', recognizing similarities and differences. Our work shows promise for designs leveraging the exchange of personalized recommendations with strangers.

CCS CONCEPTS

• **Human-centered computing** → *Empirical studies in HCI*; *Human computer interaction (HCI)*.

KEYWORDS

Self-Reflection; Content Discovery; Recommender System; Social Comparison; Filter Bubble; Persona; YouTube

1 INTRODUCTION

Social media and content sharing platforms primarily use algorithms to individualize their feeds and content in order to increase user engagement [20, 27]. These algorithms typically work by predicting what users will be interested in based on their prior interaction history [19]. This mechanism often ends up limiting the set of content that users are likely to consume, filtering out the vast majority of content available that users could have enjoyed [20]. While the resulting recommended feed may increase user engagement, users may be trapped in a "filter bubble"—that is, isolation from alternate viewpoints—potentially limiting their choices [8, 75]. However, it is challenging for users to gain awareness of their limited content consumption with all information getting intercepted by algorithms. Users with low cognitive reflection are especially susceptible to being swayed to extreme beliefs [84]. Other research shows that while some people might be aware of the existence of such filters, people hardly take actions to counteract them by methods such as clearing one's browsing history, using a browser's "incognito" function, and clicking/liking different posts [13]. Though algorithmic improvement for diverse recommendations has been an active area of research [44, 90], it still falls short of its goal, resulting in the persistence of filters [12]. Therefore, the limitations of modern recommender systems raise the need for design interventions that can facilitate diverse content discovery, reflection, and understanding.

One potential intervention that could burst algorithmic bubbles is presenting diverse viewpoints to users [65, 74]. For example, in the case of YouTube, one way to implement this solution is to show users recommendations that others received; that is, a collection of videos that YouTube's algorithms recommended to other users of the platform. We anticipate that seeing recommendations from strangers may benefit users who are otherwise exposed to a limited set of content in the following ways. First, knowing the kinds of videos that are recommended to other users can facilitate reflection on one's own tastes and consumption behaviors through social comparison. Prior research suggests that such comparison between peers could lead to improved self-knowledge or reflection [28, 93]. Second, seeing diverse recommendations from strangers could also facilitate the discovery of new content, such as content that simply seems interesting, content that specific groups of users watch, and content that a user did not know was available. In addition, presenting a proxy of strangers like an anonymous persona with recommended videos can be effective for social comparison [4, 82].

The goal of this paper is to explore the idea of exchanging algorithmically mediated recommendations as a way to facilitate content discovery and reflection, and to assess the potential barriers to such an approach in content consumption. We accomplish

this by designing, developing, and evaluating OtherTube—a browser plug-in for YouTube—which records the videos recommended to a user from the YouTube homepage and displays them to others. OtherTube allows users to see strangers' YouTube recommendations (see Figure 1) as part of the homepage. To facilitate better social comparison, OtherTube also lets users create an anonymous persona (see Figure 3) and display it alongside their recommended videos. Furthermore, OtherTube lets a user remove recommendations that they do not want to share. Using our design, we aim to answer the following research questions:

RQ1. *How do users discover content by browsing recommendations personalized for strangers?*

RQ2. *What factors affect users' interactions and engagement with recommendations personalized for strangers?*

RQ3. *How does browsing recommendations personalized for strangers facilitate reflection?*

To answer them, we conducted a 10-day long user study with the plug-in. There were 41 participants in the study; participants had to use and interact with OtherTube each day and fill out a daily survey. In addition, we logged information about the interactions that occurred within OtherTube during that period, such as the number of videos clicked and the number of clicks to see different personas. To better understand participants' behavior during the study, we conducted semi-structured interviews with 11 participants.

Our analyses show that OtherTube can help some users, but not all, to develop new interests and rediscover old ones by seeing strangers' personalized recommendations. We also found that factors such as lower age and a lower need for self-reflection had a positive impact on the extent to which users browsed OtherTube content. After viewing others' recommended videos, users understood more about their interests and how unique those interests were. Encountering other users with similar interests also gave users a sense of belonging. Overall, our results inform future developers of online social systems of various design considerations for allowing the exchange of recommendations: supporting content discovery and reflection needs, and encouraging cross-demographic interactions. We conclude by exploring the implications of our approach in creating content recommendations, curating users' interests, and building social connections.

2 RELATED WORK

In this section, we briefly review existing research around content discovery and reflection pertaining to recommender systems. With our system supporting comparison with strangers, we also review related literature.

2.1 Supporting Diverse Content Discovery Online Through Recommendations

Social recommender systems have become ubiquitous over the last decade, in areas such as social media (e.g., Facebook), e-commerce (e.g., Amazon), video sharing platforms (e.g., YouTube), and recreational services (e.g., Netflix). As recommender systems have become highly accurate in estimating users' preferences [40], it also comes with caveats when their suggestions become repetitive and monotonous [92]. With the continuation of this process, it often leads to filter bubbles where algorithms discard content opposing users' interests [75]. These problems have prompted inquiry into diversifying users' exposure to differing viewpoints [78]. Throughout this work, we use traditional definition of exposure diversity, i.e., the supply-side (source and content) diversity [43, 67]. To improve diverse exposure, one line of research takes diversification as a quality metric for recommender systems and introduces novel approaches to improve it [14, 44, 54, 61, 90]. For example, some scholars introduced recommendations based on "Social Diversity", i.e, presenting recommendations from users of different social groups who are not connected [83]. To generate diverse recommendations, other shcolars have suggested using content that are less popular among all users [1]. Still, others have used statistical models accounting for inter-item correlation to generate diverse recommendations on YouTube [90]. While analysis of these systems deployments show some measurable improvement, they are still lacking due to the trade-off between diversity and accuracy [94]. Furthermore, subsequent suggestions by recommender systems often depend on a user's prior interactions. Therefore, if a user does not interact with diverse recommendation, the following ones may not be so diverse [68]. This issue can explain the persistent lack of diversity in most recommender system outputs, despite attempts to addressing diversity; and, our design does not depend on user interaction.

In parallel, there has also been some research over design-centric approach to address the issue of filter bubble on social platforms [34, 52, 62, 73, 81]. These approaches include design interventions to understand users' own content consumption habit by showing information such as their topic-wise content consumption [81], political leaning of the sources they consume information from [65], credibility of the content they consume [9, 10] and political leaning of their own social network [34]. Some of these approaches also promote viewpoints from alternate perspectives, such as, related content from alternate sources [73] and viewpoints from a user with different political ideology [52]. Our work adapts the approach of showing alternate viewpoints for YouTube by extending a particular demography-based feed exchanging approach to a stranger-centered feed exchanging one.

2.2 Designing for Reflection Online Using Recommendations

Reflection has received significant attention in HCI works. Existing research has shown promise of reflection in various areas such as education [36, 46, 53, 86], health or wellbeing [30, 39, 80, 85], and self-knowledge or personal informatics [3, 24, 55, 57, 58]. Several models have been proposed around personal informatics systems including Li et. al.'s Stage-Based Model of Personal Informatics Systems [57], Epstein et al.'s the Lived Informatics Model of Personal Informatics [24] and Niess et. al.'s Tracker Goal Evolution Model [70]. These models define reflection as a set of stages, from motivation to reflection [24]. Eventual purpose of many of these systems is to facilitate behavior change [7, 18, 60], especially through goal-setting [51, 63]. In the past, designing for reflection around online social space has also received some attention [5, 21, 56]. Some of the prior work on social media revolves around personal informatics [21, 56]. For example, several works track affective expressions in users social activities (e.g., posts and comments [21, 48]), thus providing insight into their personal behavior.

Literature shows that many systems for reflection and desirable behavior changes—e.g., eating healthily, regular exercise, enhancing productivity—relies on self-tracking. However, realizing one's filter bubbles further requires elements of social comparison where such comparison can help users recognize what is missing. Literature terms this method as *interpersonal informatics* [6]. Works on *interpersonal informatics* suggest that it can be effective for users to understand their position through projecting oneself onto others. For example, Feustel et. al. examines reflection using cohort data from multiple sources [29]. Different properties of data may also stimulate reflection including showing invisible information, allowing to compare, revealing ambiguity and providing multiple perspectives [64]. For example, some previous work shows promise of providing data for comparison [87]. However, use of recommendations in this space is sparse. Among the existing works, some used personalized recommendation as a tool to trigger reflection on a particular artifact [49, 71]. Some systems also used visualization of unfiltered and curated feeds to improve users understanding of recommender systems by reflecting on them [25, 26]. In one sense, our work promotes interpersonal informatics in the absence of social ties, by providing data from strangers. Though there are some HCI works using strangers in their designs, they relate to neither content discovery nor reflection [38, 69]. Our work draws ideas from this social comparison used for reflection and employ the approach of exchanging recommended content.

2.3 Enabling Social Comparison for Reflection

The theory behind behavior change leveraging social comparison is not new [28]. In some cases, such comparison could act as a support. In others, social comparison can trigger peer pressure which promotes competition [16, 76]. Prior studies found this mixed effect within the same system [17, 59, 91]. Research also shows that constructing better self presentation for social comparison on sites like Facebook may lead to improved self concept and self-esteem [32, 93]. Motivated by these positive and negative implications of social comparison, we designed OtherTube to allow users to create an anonymous persona where they can share unidentifiable information selectively (e.g., gender, race/ethnicity and age group). While comparison might be easier when users can identify the person they are comparing with, it also conflicts with users' need to preserve privacy for certain information [66]. To address this concern, Garbett and colleagues used pseudonyms and avatars, protecting users' identities in facilitating self reflection in a group setting [31]. For YouTube, with its long history of toxic interactions [15, 72], we use a similar approach to anonymize users' self-presentation (using pseudonyms, avatars and generic demographic information) in our design of OtherTube.

3 OTHERTUBE: DESIGN & IMPLEMENTATION

To provide an environment that can be integrated into users' YouTube usage, we built OtherTube. OtherTube is a Chrome extension usable across all operating systems; users need only use the Chrome browser to browse YouTube. Our system works by collecting a user's YouTube recommendations—specifically, the top two or three rows of videos—each time a user visits the YouTube homepage. It stores the recommended videos in a database to be shared with

strangers from the next day onward. In short, users are given access to strangers' recommended videos (see Figure 1- 4) in exchange for providing their own recommended videos to strangers. Figure 2 demonstrates this process. Additionally, OtherTube provides three main affordances: (a) an option to allow users to create an anonymous profile (Figure 3), (b) an option to remove collected recommendations that they may not want to share (Figure 4), and (c) an option to choose between browsing strangers' profiles[1] and recommendations from users' own YouTube homepages (Figure 1). We describe each of these affordances below.

3.1 Creating an Anonymous Profile

To give users extra information about the strangers whose YouTube recommendations they are browsing, OtherTube asks each user to create an anonymous profile. For each user, we generate a random screen name—a combination of an adjective, a noun, and a number (e.g., amazedOtter4)—automatically in the back end. Users can choose an avatar for their profile and set several demographic attributes. Figure 3 shows the available options. The avatar builder allows users to choose a skin tone, hair color, clothing color, and appearance[2]. For demographic attributes, users have the option to share or not share their age, gender / gender identity, ethnicity, race, and location. To provide anonymity, users can only set their age as a decade-based bucket (e.g., Teen or Twenties) and location by state or province. Users also have the option to enter their own values for certain attributes. Apart from the demographic details, users can choose to fill out an open-ended "About Me" section in their profile. For example, Figure 1 shows a profile with the text "Love traveling, food, and my family" in this section. At the top of the page and the About Me input field, we put disclaimers asking users not to share any personally identifiable details. Users can update their profiles at any time. Finally, note that OtherTube does not start collecting recommendations from a user or showing recommendations to others until users have created a profile. This ensures that users only see collected videos attached to profiles.

3.2 Sharing and Removing YouTube Recommendations

Each time users visit the YouTube homepage, OtherTube collects their YouTube recommendations and sends them to a back-end server, which then stores them in a database. Going forward, we will call each visit a *session*. It is worth noting that the recommended videos are generated by YouTube's algorithm; they do not simply consist of a user's browsing history. This means that the collected videos do not constitute an interaction trace, so the plug-in does not have to monitor a user's entire watch history, which could be perceived as private data. Once a session of recommended videos is stored on the server, the recommendations are shared with other users over the following days. If users do not want to share certain YouTube recommendations with strangers, they can choose to remove individual videos from the set collected by OtherTube. When users click on the extension button next to Chrome's address bar, the extension shows a list of sessions sorted by time, with the most recently collected items at the top. While browsing the collected

---

[1] Throughout the text, we use the terms *persona* and *profile* interchangeably.
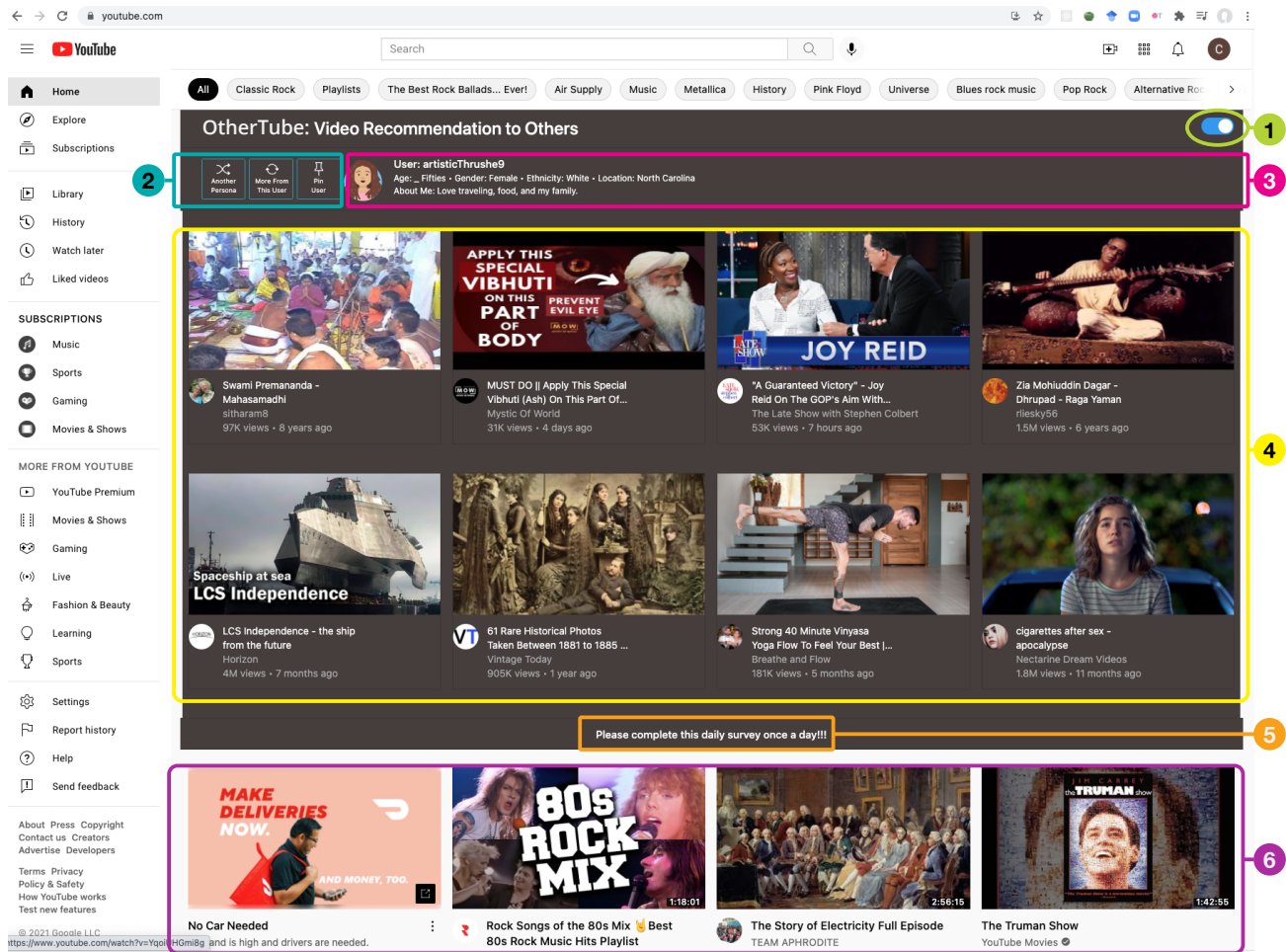[2] We used a third-party library, *AvataaarsJs*, for the avatar builder.

Figure 1: OtherTube embedded inside the YouTube homepage. ① Show or hide the embedded content. ② Browse different strangers or different recommendation sessions from the current stranger, and pin the current stranger. ③ The stranger's profile. ④ YouTube recommendations collected from this stranger. ⑤ Link to a daily survey. ⑥ The user's own YouTube recommendations, which OtherTube collects.

sessions, users can click a "remove" button in the upper-right corner of each video (the blue trash can icon in Figure 4) to remove content that they would rather not share with strangers.

3.3 Browsing Strangers' YouTube Recommendations

OtherTube embeds recommendations from strangers, which are collected through the steps described in Sections 3.1 and 3.2, at the top of the YouTube homepage (see Figure 1). OtherTube differentiates itself from users' own YouTube recommendations on the homepage using a brown background color. Users can hide or show recommendations from OtherTube using a toggle in the upper-right corner of the embedded content (Figure 1-①). In the upper-left corner, OtherTube displays three buttons that allow users to browse strangers' recommendations ("Another Persona"), browse another recommendation session within the same stranger's persona ("More

From This User"), and pin this stranger's persona ("Pin User") (Figure 1-②). Each time users click Another Persona or More From This User, the server returns a random stranger's recommended videos or a random session from the current stranger, respectively. When a user clicks Pin User, a shortcut to the displayed user's profile and recommended video collections is created below the button. We added this button in case a user wants to follow and revisit a particular stranger's recommended videos. To the right of the three buttons, the current stranger's profile is shown, consisting of their avatar, their demographic info, and their About Me text. Below the profile, OtherTube shows the stranger's YouTube recommendations. Below the recommended videos, there is a link to a daily survey which we asked participants to fill out each day during our study (see Figure 1). Finally, OtherTube only tracks users' interactions within the plug-in (e.g., clicks on the Another Persona, More From This User, and Pin User buttons; and clicks on videos).
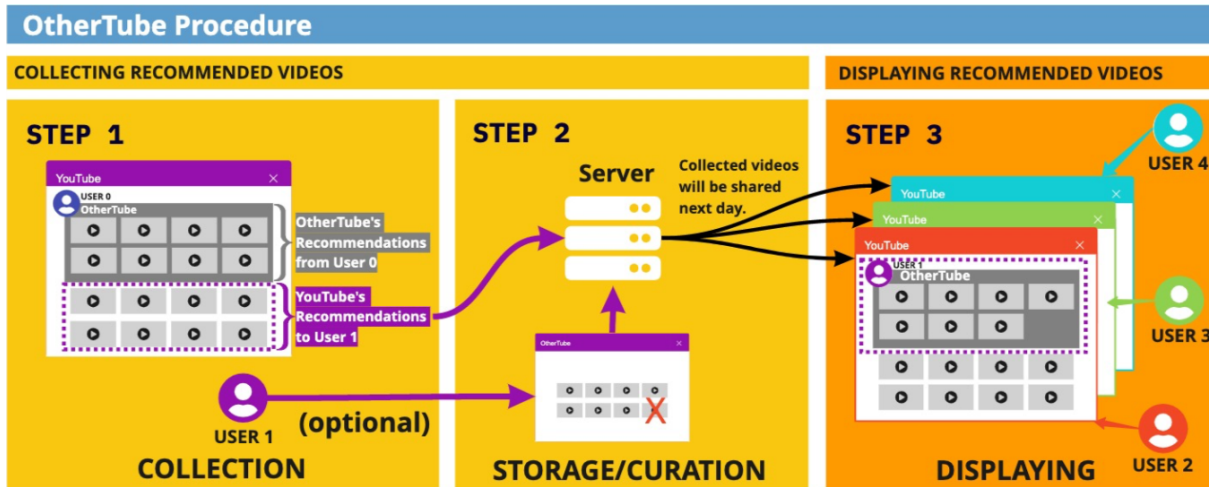
Figure 2: How OtherTube works. Each day, OtherTube collects YouTube recommendations when users access the YouTube homepage. Users have until the end of the day to remove items they do not want to share. Users can browse recommendations collected from others as recently as the previous day.

Figure 3: OtherTube's Options page. **1** The avatar builder. **2** Shared demographic information. **3** How the user's profile will appear to others.

Figure 4: OtherTube Browser action page. **1** Collected videos with options to remove from the shared set.

We built the front end of OtherTube using the React and Polymer JavaScript libraries, with Bootstrap CSS for styling. The back end consists of a Flask-Nginx server with a MySQL database for storage. All communication between the front end and back end is encrypted using SSL. After building the tool, we tested it within our research groups and ran a pilot study, fixing technical issues and improving usability. We distributed OtherTube through the Chrome Web Store.

4 STUDY DEPLOYMENT

Using OtherTube, we conducted a 10-day-long study. This study was approved by the university's Institutional Review Board. Below,

we outline our recruitment method, study procedure, data collection process, and analysis.

4.1 Recruitment

For our study, we aimed to recruit participants who use YouTube on a regular basis. In addition, we decided to recruit participants using social media, specifically using Facebook Ads, informed by others' successes in recruiting diverse populations [2, 77, 89]. This advertising technique allowed us to reach a more diverse and targeted demography compared to Amazon Mechanical Turk or dedicated survey sites like Qualtrics [11]. Initially, we ran an advertisement campaign targeting individuals living in the US who are 18 years of age or older, speak English, and are interested in YouTube videos. While limiting our target demographic to those who live in the US would limit our findings, we did not want to have to account for language barriers in exchanging video recommendations. Our

goal was to still reach diverse populations in terms of age, gender, and ethnicity. In addition, studying users living in a single nation provides some useful common ground upon which they can relate their interests to those of others (e.g., popular artists and domestic news items). The recruitment campaign was set to run for one week, from July 8, 2021 to July 15, 2021. We spent $315 on the campaign and received 568 responses.

4.2 Procedure

Users who clicked the Facebook advertisement were redirected to a pre-survey to sign up for the study. At the beginning of the pre-survey, we screened users according to several criteria. To be eligible for the study, users had to (i) be 18 or over, (ii) currently reside in the United States, (iii) visit YouTube at least once a day, (iv) typically browse YouTube on a laptop or desktop computer (as opposed to mobile-only users), (v) typically use Chrome to browse YouTube, (vi) have English as the primary language of the YouTube content they watch, and (vii) typically start browsing YouTube from the YouTube homepage (youtube.com). Out of the users who submitted the presurvey (n = 568), 318 were eligible for the study. We invited these participants via email to complete the consent form and start the study[3]. The invitations were sent out in two batches: one from July 19 through August 3, 2021, and another from August 5 through August 18, 2021. Note that while the emails to all participants in a given batch were sent out on the same day, users could have started using the extension on different days, resulting in batch periods exceeding 10 days. Out of our 318 invitees, 41 participated in the study by installing the plug-in and filling out the daily survey at least once. We created an instructional document describing how to participate in the study. Users who accepted the invitation had to install the OtherTube extension from the Chrome Web Store. After installing the extension, users had to create profiles. At the beginning of the study, to mitigate the cold start problem, we created a research account so that participants could begin to see embedded recommendations from the first day. For 10 consecutive days, users were asked to use YouTube as they normally would and interact with OtherTube. Each day, they were also asked to submit a daily survey which took about five minutes to complete. We sent reminder emails around 6 P.M. EDT each day to remind users who had not yet submitted the daily survey. Despite the reminders, participants did not consistently submit the survey, leaving us with 356 (8.7 on average) responses instead of 410 (41 participants × 10 days). Upon completion of the study, we invited about half (n = 19) of the participants to an interview based on their survey completion rates. Of those invited, 12 participated in the interviews. Each meeting was recorded for analysis. Because one user revealed that they neither followed the study instructions nor had a clear understanding of how the plug-in works, which would have left the user with no context for many of our questions, we ignored this user's response and analyzed the remaining 11 recordings. We compensated study participants with $25 gift cards and interview participants with $15 gift cards, in line with federal minimum wage requirements.

---

[3]Initially, we prioritized minorities for invites to form a diverse pool. Due to the limited response, we eventually reached out to all participants.

4.3 Participants

Our pre-study survey was mainly designed to filter out ineligible users and create a participant pool that represent diverse demographic groups. However, we could not completely fulfill this objective due to an inconsistent response to study invites. Figure 5 shows the distribution of age, gender, political affiliation, and length of a typical YouTube browsing session among our 41 participants. The participants' demography is balanced in terms of gender. However, it is heavily skewed in terms of race, with only one Black or African American participant despite a sufficient number of Black or African American users signing up for the study (see the contrast in Appendix A). This disparity could be caused by hesitancy towards installing tools or hesitancy towards research studies due to past injustices [45, 50]. By political affiliation, the majority were Democrats. In the About Me section of their profiles, participants mentioned various occupations (e.g., "manager" and "recent college graduate") and interests (e.g., photography, traveling, yoga, DIY, Hololive, and BTS).

4.4 Data Collection

We collected data from our participants in multiple ways, beginning with the pre-study questionnaire. The questionnaire was followed by interaction traces and daily surveys during the study, and the post-study interviews came last.

4.4.1 Pre-Study Questionnaire: Need for Self-Reflection and Insight. In the pre-study questionnaire, along with demographic questions, we asked about users' need for self-reflection and insight (see Appendix B for the items), using scales from prior research [41] on 5-point Likert items with responses from "strongly disagree" (1) to "strongly agree" (5). We measured these metrics to see if the self-assessed need for self-reflection and insight would correlate with how users interact with OtherTube. Participants' responses had a high level of consistency for both questions, similar to prior studies (Cronbach α [Need for Self-Reflection]: 0.97, Cronbach α [Insight]: 0.95) [37, 41]. For each trait, we found the mean of the items after inverting items that were phrased in the opposing sense. Figure 6 shows the distribution of user responses for need for self-reflection and insight.

4.4.2 Interaction Traces from OtherTube. We collected data on users' interactions with OtherTube, including clicks to see recommendations from different strangers, clicks to see more from a stranger, clicks to watch videos, clicks to pin strangers, and clicks to remove own recommendations.

4.4.3 Daily Survey: Perceptions of Recommendations on OtherTube. In the daily surveys, we asked users Likert scale and open-ended questions about their experiences with OtherTube(Table 1 lists the questions). We included several Likert items on a 5-point scale from "strongly disagree" (1) to "strongly agree" (5) about users' perceptions of OtherTube on YouTube. These questions captured users' perceptions of content discovery each day, whether their own YouTube recommendations were affected by their interactions with OtherTube, and how they felt about sharing their YouTube recommendations with strangers. Through open-ended questions, we aimed to understand participants' reflections on themselves and
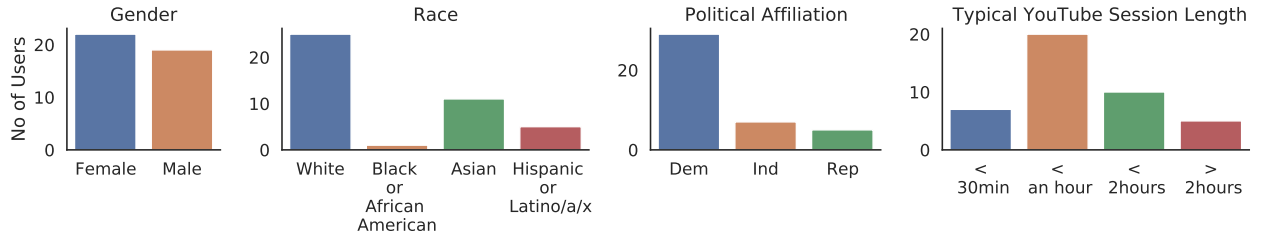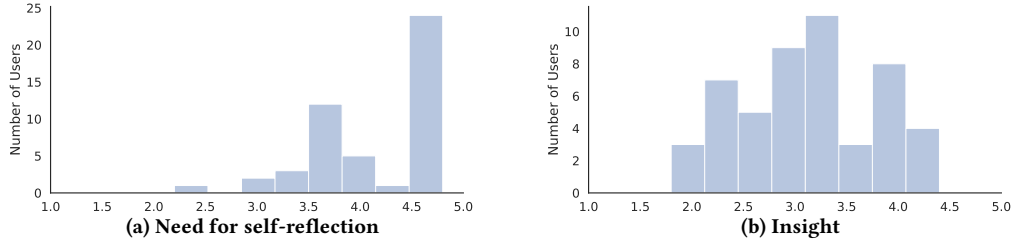
Figure 5: Demography of the participants in the study.

Figure 6: Distribution of participants' need for self-reflection and insight, bucketed for ease of understanding.

Question Type	Item
Likert-scale	(a) Today, I saw some videos on OtherTube that caught my attention.
	(b) Today, OtherTube recommended me videos that I would not have expected in my feed.
	(c) Today, I discovered some new types of content on OtherTube that I would like to watch.
	(d) Today, YouTube recommended me some videos that I would not have expected in my feed.
	(e) Because of using OtherTube, YouTube is recommending diverse videos to me.
	(f) I feel comfortable sharing my recommended videos with others on OtherTube.
Open-ended	(a) If there were any videos that caught your attention from OtherTube, could you tell us what they were and why you were interested in them?
	(b) If you removed any recommended videos of yours from the OtherTube plug-in, could you tell us why?
	(c) While using OtherTube, did you learn anything new about certain populations (different age group, different gender)? If so, what did you learn?
	(d) While using OtherTube, did you learn anything new about your own taste compared to others? If so, what did you learn?
	(e) What made you hesitate to watch videos from OtherTube, if any?

Table 1: Daily Survey Questions

others, and to learn the motivation behind any actions they performed or chose not to perform (e.g., content removal or hesitation to watch content).

4.4.4 Interview. After completion of the 10-day study, we interviewed 11 participants. The interviews focused on understanding how they used the features of the plug-in, and whether using OtherTube encouraged content discovery or facilitated any reflection. Additionally, after analyzing each interviewee's daily survey responses, we asked them to elaborate on any points that seemed unclear. For example, because one user mentioned that their interests are "conservative" in their daily survey, we asked them to elaborate on what they meant by that (i.e., whether the user was interested in political conservatism or had conservative viewing habits with preference to watch videos with limited interests). We concluded the interview with some usability questions. See Appendix C for the complete interview questionnaire.

4.5 Method of Analysis

We gathered both quantitative responses (pre-study survey, Likert scale–based daily survey questions, and interaction traces) and qualitative responses (open-ended daily survey questions and interview responses). Below, we describe our analytical methodology.

4.5.1 Likert Items on the Daily Survey. To gain insight into self-assessment on the statements, we checked the average responses to each Likert item on the daily survey. We also performed Mann-Whitney U tests, a nonparametric test, to see if using OtherTube produced any changes in their perceptions, comparing the responses from participants' first and last days of using OtherTube. Because not all users submitted the survey on each of the 10 days, the last day might not have been the 10th day for each user.

4.5.2 Interaction Traces. To answer RQ2, we modeled the number of daily clicks on the Another Persona button using negative binomial regression[4]. The independent variables for the regression consisted of users' demographics, the need for self-reflection, and insight. Because we recorded the number of clicks on each of 10 days, resulting in repeated measures for each user, we used a mixed-effects regression model. Additionally, because clicks are a count variable, we used negative binomial regression[5]. Because some users did not click the buttons every day and random effects require multiple observations per user, we used users who clicked the button on at least three days. Therefore, instead of 410, we had

---

[4]We repeated this process for the More from This User button, but the results were not significant; therefore, we omitted them from this paper.
[5]Due to overdispersion, we chose a negative binomial model over Poisson regression.

280 data points from 32 users (an average of 8.8 data points per person) for our model. We used *mixed_model* from the *GLMMadaptive* R package [79]. For the sake of interpretation, we present marginal coefficients instead of fixed effect coefficients in this model[6]. Additionally, we computed the Spearman rank correlation, a non-parametric statistic, to assess whether a correlation exists between users' interactions and engagement; that is, clicks on videos, Another Persona, and More from This User. Apart from these tests, we also analyzed other simple statistics.

4.5.3 Open-Ended Daily Survey Responses. We performed thematic analysis on the open-ended daily survey responses. With five open-ended questions, we had 1,780 responses (356 daily survey submissions × 5 questions), 976 of which were either empty or contained unhelpful responses, like "no" or an incomplete response (e.g., "Yoga videos" for Open-ended-(a) in 1). This left us with 804 valid responses. Three researchers performed thematic analysis on this data. Initially, each of the coders came up with their own set of codes. After discussion, we converged on a set of 52 codes. Through discussion, we reduced this subset of 52 codes to a list of 21 themes. Two of the researchers coded a sample (159 items, or 20%) into the set of 52 codes. Coders had almost perfect agreement despite the large number of codes (Cohen's $\kappa = 0.82$) [88]. One of the coders coded the rest of the items. Due to the uneven number of codes associated with each user, as some did not regularly fill out the daily survey, we merged the 10-day codes for each person into a single set. Consequently, as we present our results, "theme (10/41)" means that out of 41 users, 10 users' responses contained at least one response belonging to the theme. Each user may have given such a response anywhere from 1 to 10 times over 10 days.

4.5.4 Interview. Similarly to the open-ended survey responses, we performed thematic analysis on the interview responses. One of the researchers performed an initial analysis and came up with 35 codes from the interviews. Then, this researcher discussed the codes and corresponding quotes with other researchers. After resolving disagreements, we were left with 24 codes from the interviews. As most of them were related to the daily survey themes, we merged the two sets.

5 RESULTS

Based on our research questions, we present results consistent across daily survey themes, interaction trace analysis, and interview themes. Note that the daily survey and interview quotes are presented in the format "U1 (survey)" and "U1 (interview)".

5.1 RQ1: Content Discovery

In the daily survey, the majority of participants agreed with the statement (mode: 4, "somewhat agree") saying that they saw some videos on OtherTube that caught their attention. Further analysis shows that this perception did not change between the first and last days. This result is shown in Figure 7-(a), where a Mann-Whitney U test shows no significant difference. Our analysis revealed two

themes that illustrate how users utilized OtherTube to discover content: (i) users developed new interests, and (ii) users rediscovered content they used to like. We present these themes below.

5.1.1 **Users developed new interests**. In the daily survey, a majority of our participants ($n = 36/41$) mentioned finding new interests. Users found interests in multiple ways. Sometimes, they found new interests out of curiosity (*"I had no clue what it [a video that caught their attention] was about but just curious on the content"* - *U1 (survey)*). Other times, new content led to the development of a new interest due to its usefulness (*"... A video about investing for your kids' future. I have a newborn and want to do that for her"* - *U2 (survey)*). There were also cases where users found new content that fit within their existing interests (*"Relaxing Video Game Music in a Cozy Room (Nintendo 64) [caught my attention]). I enjoy similar relaxing music playlists"* - *U12 (survey)*. Some users also watched content for the sake of exploration (*"I don't normally listen to that [classical music] and it was a welcome change"* - *U17 (survey)*). Interviewees also responded similarly, with one mentioning how they bumped into new interests on OtherTube. Merely encountering the embedded interface of OtherTube could trigger changes in consumption behaviour.

> *"What I like about it is that some it forced me out of my comfort zone of what I would just do ... It makes you to stop and, like, look and think before you decide to make a different choice."* - **U17 (interview)**

We also noticed that participants were able to recognize profiles that had similar tastes to their own (*"I came across a user who had watched videos that were my interest. He was interested in computers and video games. So, I liked those videos too"* - *U20 (survey)*). This tendency demonstrates that one can find not only interesting videos using OtherTube, but also people one can connect with, due to similarities in the content that they watch. OtherTube's additional functions to retrieve more videos from a particular user can be useful to further explore a person's preferred content. Analyzing the number of clicks on More from This User and the number of clicks on videos users watched, we found a statistically significant positive correlation (Spearman's $\rho = 0.42$, $p < 0.001$). This pattern suggests that finding a profile based on similarities in taste could be a new way to find and watch new content—in essence, subscribing to other viewers, not creators, to follow their content consumption patterns. We discuss these implications further in Section 6.1.

5.1.2 **Users rediscovered content they used to like**. Using OtherTube, some users ($n = 11/41$) found content they used to like (*"old throwback videos from my childhood [caught my attention]"* - *U11 (survey)*). They also found channels they used to like (*"a video game reviewer, Zero Punctuation, popped up in one of the OtherTube recommendations. So seeing that was pretty nostalgic."* - *U3 (survey)*).

Apart from these themes, our participant interviews further revealed users' issues with finding strangers with similar interests. For example, while users' About Me details were sometimes useful, they wanted more features, such as filtering users by demographic or interest (*"Their favorite content creator, maybe your favorite video."* - *U26 (interview)*).

Summary: By browsing OtherTube, participants both found new interests and rediscovered old ones. They found new interests

---

[6]Given the nonlinear link function (*Log*) in our model, random effect intercepts can have a multiplicative effect, rather than additive, complicating interpretation. Therefore, following Hedeker et al. [42], we extracted the marginal coefficients and their standard errors from the model using the *GLMMadaptive* R package.
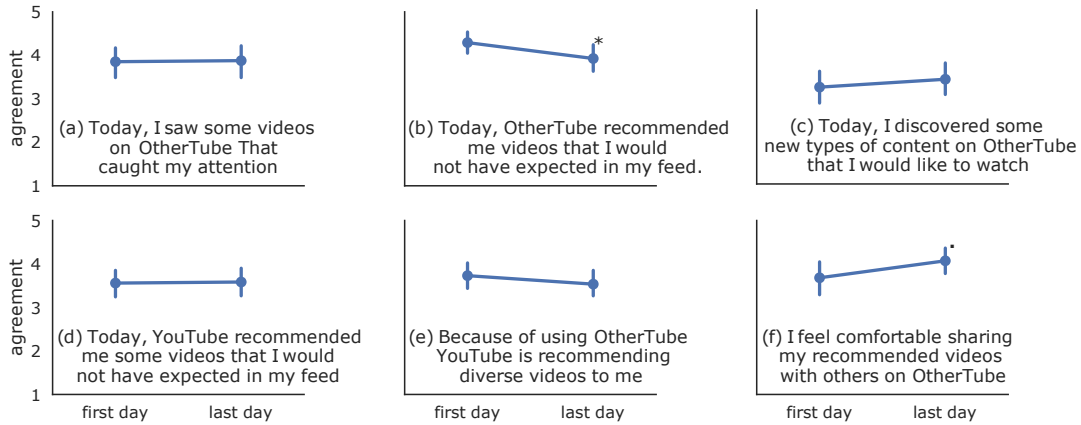
Figure 7: Mean with 95% CI of participants' responses to the daily survey Likert items from the first and last days of the study. We also performed a Mann-Whitney U test comparing responses on the first and last days. In (b), * indicates $p < 0.05$; in (f), . indicates $p < 0.10$.

Click Count per Day (Another Persona)		
	β	std. Err.
(Intercept)	5.5024***	1.0936
Age	-0.0284*	0.0135
Gender[Male]	0.2592	0.2303
Race[Black or African American]	-1.2113	0.7136
Race[Hispanic or Latino/a/x]	-0.1595	0.3895
Race[White]	0.2526	0.2412
Daily Browsing Length[less than 1 hr]	-0.6087	0.3839
Daily Browsing Length[less than 2 hrs]	-0.1536	0.3628
Daily Browsing Length[greater than 2 hrs]	0.1773	0.4109
Need for Self-Reflection	-0.4149***	0.2108
Insight	-0.0829	0.1730
Day	-0.0453*	0.0185
# of Users Available on the Day	-0.0083	0.009
Dispersion	0.9119	0.1244
log.Lik		-721.8501
N = 280	* p<.05, ** p<.01, *** p<.001	

Table 2: The mixed-effects negative binomial model for daily click count on Another Persona in Figure 1. In this model, user is a random-effect variable. User demographics, their browsing habits, their need for self-reflection, the day when the clicks were counted, and the number of unique stranger data sets available to browse on OtherTube on the day when the clicks were counted are fixed-effect variables. The estimated negative binomial regression coefficient β is the difference in the logarithms of expected counts of the response variable due to a one-unit change in the predictor variable.

out of curiosity, usefulness, or desire to explore. Items caught participants' attention similarly on both the first and final days of the study. Finally, participants also recognized others who had tastes similar to their own.

5.2 RQ2: Factors Affecting Interaction and Engagement

To investigate what factors affected users' interactions with Other-Tube, we modeled users' click activity on Another Persona over time using a mixed-effects negative binomial model. We considered several factors in this model, including users' demography, their need

for self-reflection, their YouTube browsing length, and the number of unique profiles available to browse on a given day. Table 2 shows the result. We find that older users tend to interact less with OtherTube when browsing different personas ($\beta = -0.03, p < 0.05$). Our users with a greater need for self-reflection also interacted less with OtherTube ($\beta = -0.41, p < 0.001$). This result was particularly interesting because we hypothesized that those who believe that they need reflection would use OtherTube more than those who did not have such thoughts, given our goal of facilitating reflection behind the tool. Perhaps our design will be more effective and promising for those without an explicit desire for self-reflection. Additionally, as the days passed, users interacted less with the tool ($\beta = -0.05, p < 0.05$). This decrease could have been caused by seeing the same profiles repeatedly, due to the limited number of users in the study (*"[On day 4] i haven't seen many changes. [On day 6] no same as yesterday." - U4 (survey)*). To understand users' engagement with the videos, we analyzed users' clicks on videos. We found that most users ($n = 37/41$) watched at least one video on OtherTube. Overall, participants collectively clicked on 6% or 398 videos out of all the collected videos on OtherTube.

While demographic attributes and other environmental factors can affect interaction, it might have been the content itself that encouraged or discouraged interaction with content from OtherTube. Analysis of users' daily survey responses revealed content-related factors that discouraged users from watching videos; these are discussed below.

5.2.1 ***Users hesitated to watch content that was not of interest to them.*** A majority of our users ($n = 25/41$) mentioned that they did not watch content because the content did not relate to their interests (*"I didn't find them [videos I hesitated to watch] interesting" - U40 (survey)*). Some users referred to such content as "boring" (*"seemed boring or predictable" - U35 (survey)*) or "complicated" (*"looked boring, or too techy and complicated" - U36 (survey)*). In the interview, one of the participants further elaborated on how their motivation for browsing YouTube at times affects whether they would engage with certain type of content (*"Sometimes I don't*

want to be educated. I just want to enjoy. Other people want maybe to learn about the history of Macedonia or something. I don't sometimes want to learn that." - U15 (interview)).

*5.2.2 **Users hesitated to watch content that was not helpful**.* In contrast, content utility was the main factor for some users' ($n = 6/41$) engagement with content in OtherTube (*"I just didn't click the ones that seemed pointless and not helpful to my personal development or learning" - U2 (survey)*). Some expressed hesitation in terms of wasted time (*"They could be a waste of time to watch if I don't like them" - U26 (survey)*).

*5.2.3 **Users hesitated to watch disturbing and offensive content**.* Some users ($n = 7/41$) were also hesitant to watch content that seemed disturbing or offensive to them. Various types of content could fall into this category, including horror videos (*"I didn't want to watch horror" - U13 (survey)*), sexually charged videos (*"I didn't want to watch videos that seemed strangely sexually charged, like mouth ASMRs" - U38 (survey)*) and uncomfortable topics (*"I was disturbed by videos in uncomfortable topics I wouldn't want to think about, such as, physiological anomalies and unsolved murder mysteries" - U39 (survey)*).

*5.2.4 **Users hesitated to watch clickbait-esque content**.* Clickbait was one reason some users supplied for not watching content ($n = 4/41$). Users used both video titles (*"Titles were terrible and looked like trash" - U16 (survey)*) and thumbnails (*"I was hesitant to watch videos with objectifying thumbnail pictures of women" - U38 (survey)*) to rule out this kind of content.

Some of the barriers discussed so far may reduce the overall effectiveness of our approach, as users may consider the content displayed by OtherTube to consist of low-quality (e.g., spam, clickbait-esque videos) or even explicit (e.g., profanity) items. Future designs could introduce some mechanisms (e.g., automated spam filtering) to mitigate such perceptions.

*5.2.5 **Users hesitated to watch videos for fear of an interaction effect on their YouTube recommendations**.* A few of our participants ($n = 4/41$) were hesitant to watch content because they assumed doing so would affect their future YouTube recommendations (*"I saw an interesting video on Netscape, but didn't click given it's from a new channel I'm not familiar with, but also being aware that the YT algorithm is going to try and keep me engaged by literally suggesting more of said video" - U25 (survey)*). Some of our interviewees further revealed that they deal with content that they do not want to watch by blocking it. Similar blocking mechanisms could be added to OtherTube to filter out content that users do not want to see. We leave this for future exploration.

> *"I curated it [YouTube] fairly carefully. Occasionally I would get some random stuff [on YouTube recommendations]. Usually I just click on the little dots [a button on YouTube videos with a drop-down menu to block content users do not want to be recommended] and I'd say 'don't recommend this channel'."*
> - **U39 (interview)**

This issue raises an interesting question of how users' engagement from watching content on OtherTube could affect their future recommendations from YouTube. In light of this, should designers make OtherTube available within a sandbox or incognito viewing mode? Sandboxing may encourage some users concerned about

impacting the recommendation algorithm to interact and explore more content.

Summary: Our analyses show that factors such as lower age and lower need for self-reflection positively affected users' daily interaction with OtherTube. However, participants' interaction also subsided as days passed. We additionally found that 90% of our participants watched at least one video on OtherTube. Some content factors also discouraged engagement, including lack of relevance to users' interests, lack of utility, and offensive or clickbait-esque nature.

5.3 RQ3-1: Self-Reflection

Using OtherTube, our participants were able to reflect on themselves from various perspectives. We found the following themes here: (i) understanding of own interests and their uniqueness, (ii) feelings of belonging from seeing users with similar interests, and (iii) feelings of superiority from comparing content. These patterns help us understand the kinds of self-reflection that the system can facilitate. We illustrate these themes below.

*5.3.1 **Users understood more about their own interests**.* One of the effects of using OtherTube seems to be users ($n = 14/41$) understanding their own interests, what they like (*"I learn that I have interest in cooking after watching some cooking videos" - U41 (survey)*), and what they do not like (*"I am not interested in animation or anything about computers or how computers work" - U15 (survey)*). Some also realized their preferences in terms of video length (*"I like short and fun videos, nothing too long or too serious." - U16 (survey)*). Others realized that their identity influences their interests (*"I saw my tastes to be very related to my identify [sic] and cultural background" - U29 (survey)*).

After seeing strangers' YouTube recommendations, a majority of our participants ($n = 29/41$) further realized that their interests were unique compared to those of others (*"I learned how unique and distinct my taste and preferences are" - U9 (survey)*). Some thought that their interests on YouTube were very narrow or selective compared to those of others. Some of them defined this narrowness in terms of topics of interest (*"I mainly focus just on gaming while other people hit a lot more genres" - U26 (survey)*). These quotes show users gaining awareness about their limited interests when they can see beyond YouTube's filtered video suggestions.

*5.3.2 **Seeing others with similar interests gave users feelings of belonging**.* Users were surprised to find strangers with similar interests (*"I was surprised to see another person also to be having similar tastes as me, it was like seeing my own watch history" - U29 (survey)*). Seeing others with interests similar to their own, many of them ($n = 22/41$) realized that they are not so different from others (*"I continue to learn that we aren't so different. I have always thought that from a gender perspective, but age is really where I feel my eyes are being opened." - U5 (survey)*). Some found commonality among similar demographics (*"People near my age enjoy similar content" - U19 (survey)*). Others found commonality with different demographics (*"I like the same videos as a woman in her 50's which made me laugh because I'm 28" - U2 (survey)*). Some also expressed their alignment in interests with particular demographics over others (*"I learned that I am more likely to share interests with older*

women than men of my age" - U13 (survey)). Some users expressed both their uniqueness and their differences as part of being normal (*"I think that I am fairly normal or that weird is normal. Everyone has their own tastes." - U5 (survey)*). To see more from strangers with similar interests, OtherTube provided users with a pin button that allowed them to save other profiles for later perusal. We found that some of the participants ($n = 8/41$) indeed pinned some strangers ($n = 18/41$).

Our interviews further revealed that upon seeing users with similar interests, a few were interested in communicating with them, potentially to recommend them other interests (*"I think it'd be interesting to maybe even have that kind of connection where you're sharing. I don't want to say like you're friending someone on Facebook. But maybe sharing of content ... kind of being able to make a recommendation saying if you like this, maybe you like that." - U17 (interview)*). However, at the same time, others were not interested in communicating for fear of toxic interactions. The quote below illustrates this issue.

> *"But I've been on YouTube for years ... my impression is that overall, YouTube is an extremely toxic place. And the only way I managed to avoid the toxicity is by blocking creators who I find to be like bad people and not reading the comments section. Some of the most innocuous videos will have thumbs down on them for like, I don't know what reason. So I don't know if I want to connect with those people [who watch such videos]... And I thought that the way that you guys were doing it by making it impersonal was kind of good. Because then nobody could spew profanities at each other."* - **U39 (interview)**

5.3.3 Some users felt superior when comparing their content to others'.
Although not many, a few of our participants ($n = 6/41$) also reported feelings of superiority after seeing others' recommendations (*"Apparently, my taste is better than most people's." - U16 (survey)*). Some of them attributed this sense of superiority to the lack of variety in others' recommendations (*"I have more tastes than this lady. I seek out more content." - U19 (survey)*). Manifestations of superiority go against our goal of understanding others and could be detrimental (*"younger crowds listen to stupid stuff" - U6 (survey)*).

Summary: Using OtherTube, participants reflected on themselves and their interests. When participants found others with interests matching their own, they experienced a feeling of belonging, and vice versa. Comparing interests with others, some also felt superior to others.

5.4 RQ3-2: Understanding of Others
After seeing strangers' recommendations, participants expressed what they learned and how OtherTube increased their understanding of others. Some were surprised by the content others watch, while others were surprised by the fact that their preconceived notions about certain demographics did not always match. Users also discovered diversity in strangers' recommendations. We illustrate these themes below.

5.4.1 Users were surprised by some of the content others watch.
When browsing their OtherTube feeds, some users ($n = 11/41$) were surprised to see certain content. Sometimes, users were surprised by content when it came from a particular profile or

demographic they did not expect it to come from (*"I found it very cool that someone in their 40s was still watching Olivia Rodrigo. Makes me feel better about aging haha" - U5 (survey)*). However, comparing users' responses over time, this surprise seems to have gradually subsided. Indeed, Figure 7(b) shows a decrease in perceptions of seeing unexpected content on OtherTube, and this difference was statistically significant ($p < 0.05$). This result may indicate that participants got used to the recommended videos from OtherTube. However, it is worth noting that it is unknown whether this trend would have arisen even with users being exposed to completely new sets of profiles every day; the limited number of profiles available could have contributed to this trend. Still, the mean of the responses in Figure 7(b) did not fall below "somewhat agree" (4) on users' final days using OtherTube.

5.4.2 Users were surprised to see that certain stereotypes do not reflect reality.
A few of our participants ($n = 6/41$) were surprised to see that some stereotypes did not match reality (*"... It [different interests among people] challenges some stereotypes as I see things on their feed that I wouldn't have expected" - U40 (survey)*). Some found users of certain ages watching videos they would not expect them to watch (*"I continue to be surprised by music recommendations in particular. A lot of older persons getting younger artists and some younger persons recommended classic rock/pop" - U5 (survey)*). Some also were surprised in terms of gender stereotypes (*"I saw a male user watching some homey vlogs and that was a bit surprising" - U8 (survey)*).

5.4.3 Users learned something particular about a person or a demographic, which could be stereotypical.
Using OtherTube, about half of our participants ($n = 17/41$) mentioned learning about others' lives by watching the recommendations that strangers received (*"I learnt how you can tell about what is happening in somebody's life by looking at the videos they are watching." - U29 (survey)*). Some participants discovered what a particular demographic might be interested in (*"Men may be more into horror" - U13 (survey)* and *"40s and older watch more health related videos" - U41 (survey)*). While some people experienced a feeling of superiority when learning about others' interests, others had the opposite reaction (*"I learned to respect other's choices" - U29 (survey)*).

Meanwhile, some users ($n = 5/41$) also found that stereotypes about the kinds of content particular demographics prefer matched what they saw (*"... It reaffirms some stereotypes that I had, as some videos are expected" - U40 (survey)*). Users most often mentioned stereotypes about age, particularly for younger people (*"Today was much more in line with expectations. A teenager recommended some Pokemon videos, 20s getting anime." - U5 (survey)*). Some noticed similar recommendations related to ethnicity (*"I learned than ethnic Americans likely have video interests in their ethnic culture" - U38 (survey)*). Some also employed stereotypes to infer strangers' ages (*"Despite keeping their age hidden, my first profile today was clearly a kid/teen" - U5 (survey)*). One obstacle in understanding more about others could have been that some users kept their demographic information hidden, as allowed by OtherTube (*"several people didn't put their demographics. So I don't know their age group, gender." - U41 (survey)*).

*5.4.4 **Users discovered diversity in strangers' interests***. Some of the participants (*n* = 10/41) liked the diversity of the content they found on OtherTube (*"I like the diversity of participants and the content of the videos" - U11 (survey)*). They found diversity within particular strangers' feeds (*"... even within personas, it's getting harder to pin down a common thread. The recommendations can be pretty diverse." - U3 (survey)*). Users also found diversity across demographics (*"Every age group is different and they all post different material on here." - U4 (survey)*).

Apart from reflecting on self and learning about others, nearly half of our participants (*n* = 18/41) learned something about YouTube content broadly or what is trending on YouTube using OtherTube. Some learned about common interests of people on YouTube (*"They [OtherTube users] all have fun videos" - U11 (survey)*). Others learned of new phenomena on YouTube (*"People are using YouTube as a news source with more regularity" - U18 (survey)*). Some discovered the popularity of certain content genres on YouTube (*"K-pop is massively popular. even more than I imagined" - U17 (survey)*). Some also learned about the popularity of content they already watch on YouTube (*"Realized some content I watch is actually popular." - U25 (survey)*), or about how the YouTube algorithm can affect their interests (*"When I don't have my own tastes in videos, Youtube shapes my tastes" - U13 (survey)*).

Summary: After using OtherTube for 10 days, participants learned several things about others. Some of these things surprised users, particularly when their preconceived ideas did not match what they learned. Participants realized how diverse people's interests can be on YouTube.

6 DISCUSSION

This study presents the utility of exchanging YouTube recommendations with strangers. RQ1 reveals unique ways in which this can facilitate content discovery. RQ2 identifies demographics that might be interested in browsing recommendations catered to strangers and watching videos from them. Finally, we saw how users reflected on themselves and others through RQ3.

6.1 RQ1 & RQ2: Content Discovery, Interaction, and Engagement

While YouTube provides some features for finding new content (e.g., search or subscription, exploring trending videos), diversification of YouTube recommendations is still challenging. It is especially hard to recommend diverse content to users based solely on their watch history. As one of our participants noted,

> *"I sometimes want to find new songs ... it [YouTube] just kind of recycles a bunch of songs that I've heard ... In that instance, I actually have to basically open an incognito tab and pretend I'm just a random nobody to get actually novel or interesting music recommendations ... If I were to look for something new, I probably wouldn't be using my existing Google account essentially, rather try pretending to be a new person."* - **U25 (interview)**

In a way, our design emulates what this user has to do to get novel recommendations. Use of OtherTube shows some promising results, with participants in the study both developing new interests and rediscovering prior interests after seeing content from strangers. As

our extension embeds diverse recommendations from the content available on YouTube, it could also act as a nudge for people to find new interests. At the same time, our results also illuminate limits in the types of diverse content that can be shown. For example, recommending offensive or excessively unfamiliar content could make people more hesitant to watch new kinds of content. Consequently, there is room for improvement in designing systems for exchanging recommendations with strangers, especially in assisting users to find people with similar interests or showing profiles not completely randomly, but based on a few similarities, to prevent adverse reactions from users. We could devise intelligent algorithms based on users' profiles to show content they may like. Additionally, we could computationally filter out spam content, like clickbait videos. Others may need more case-by-case input (e.g., content on unfamiliar topics). Based on our RQ1 results, future work could introduce two browsing modes—one for developing new interests and another for discovering old interests—to facilitate better content discovery. In each case, the system could provide suggestions based on their specific goal. For example, when browsing for old interests, users can specify time window that the system could use to search recommended videos from that period. Furthermore, users' lack of interaction could also be used to improve future suggestions. For example, when people choose not to interact with a stranger's profile, profiles that are similar to the profile could be assigned lower rankings, while dissimilar profiles could be given priority for later suggestions.

Our study was devised to test how users may react to a system for exchanging recommendations with strangers. Therefore, users' reduced engagement over time is understandable in light of the small number of profiles available to browse, as seen in RQ2. Nevertheless, it remains unclear how users would react with a larger pool of participants. New problems may arise in such an environment. For example, effectively browsing or navigating through a large number of strangers until they find content that interests them can be an overwhelming task for users, especially in the long run. Perhaps new navigational techniques can be provided to the users to alleviate this issue. For example, if we allow users to filter strangers by various criteria (e.g., music interest and live video watching habit), it would be easier to browse within the group that fall under those criteria. Overall, our work lays a foundation in this domain for exchanging recommendations between strangers, and further exploration is still needed.

6.2 RQ3: Understanding of Oneself and Others

As our results suggest, OtherTube helps users understand more about their interests and compare those interests with others' on YouTube. This understanding could positively impact users if they experience a feeling of belonging by seeing others with similar interests. Then again, the opposite can also happen; that is, if users are not able to find others with similar interests, they may instead feel disconnected, isolated, or atypical. A potential solution to address the opposite effect is to design the system to suggest recommendations from users with some common ground (e.g., someone from the same demographic group, someone with similar taste) and monitor their interactions (e.g., video clicks). However, pairings of extremely similar users should be avoided, as this may lead users back into

their filter bubbles. Understanding the right amount of similarity and dissimilarity from which users can easily relate to others while still learning something is an interesting challenge. Being able to to search or filter users by demography (e.g., age range) and interests (e.g., favorite videos and content creators), as opposed to randomly suggesting profiles, may be another option for improving users' experiences with OtherTube.

Our results also suggest potential in creating social connections between strangers with common interests. Our feature allowing users to pin strangers already act as a "subscribe to another viewer" type of function. One future improvement of this feature could be helping user realize how recommendation from their subscriptions change over time. To extend the subscribe feature to a two-way connection, future work would need to consider the case of opposing desires (users who want to connect vs. those who do not). Perhaps designers can provide limited options for connecting pairs of strangers. For example, the system could unlock messaging options between two users only if they have multiple shared interests and want to connect [23]. Alternatively, communication can be centered around content, such as by allowing users to recommend videos directly and respond to recommendations in a minimal fashion (e.g., using an emoji). In addition to making connections, one potential direction for research lies in designing interventions to nudge self-reflection. Users' activities, including browsing and watching patterns, can be used to trigger such interventions. For example, questions can be devised to help users reflect on changes in their browsing habits, reflect on connections with strangers, and encourage empathy-building with others. Or, notifications can be posted to users summarizing their usage information, such as, the number of strangers' profiles visited and the number of videos watched. Overall, future works could explore these directions for supporting reflection by exchanging recommendations with strangers.

6.3   Design Implications

On platforms such as YouTube, where recommendation algorithms contribute to shaping users' tastes and potentially trap users in filter bubbles for the sake of engagement, our approach of exchanging recommendations shows new ways to expand the scope of content discovery and improve reflection. For content discovery, existing systems already adopt multiple approaches to provide users alternative recommendations to explore. For example, apart from the site's personalized recommendations, *Trending on YouTube* aims to promote exploration into "videos that a wide range of viewers would find interesting" [35]. Meanwhile, services like Spotify suggest content through features like *Discover Weekly*, based on the interests of other users with preferences similar to one's own [33]. In contrast, our approach does not aggregate interests from any particular group; rather, it promotes exploration into other individual users' interests. Because we use other individuals' YouTube recommendations, our approach is fundamentally unlike YouTube's trending recommendations. As companies such as Alphabet, which owns YouTube, become keen to provide users with more opportunities to curate and expand their interests [47], our approach may provide such an opportunity. In particular, this design creates opportunities for content consumers to follow their fellow consumers, instead of following only content creators. Aside from this,

our approach also has the potential to fill a gap in the social functionality of systems like YouTube by creating connections between strangers based on similarity in interests. Here, following strangers' recommendations can already act as a channel of communication, like a feedthrough mechanism [22]. On top of that, we may need to devise new approaches for communication. Furthermore, by using following-follower connections as a weak form of social tie, recommender systems can learn more about users by examining who watches whose recommendations and use such information to improve recommendations. Our design can also counteract filter bubbles in other settings, unlike solutions catering to only political echo chamber settings [65]. For example, merchants like Amazon can swap recommendations to promote new content to a user.

6.4   Ethical Considerations

We took several steps in our design and our study to minimize potential harm. First, we collected only data that was needed for the tool and study. These included recommendations from users' YouTube homepages and their activity within the extension: the embedded content on the YouTube homepage, the extension's browser action page, and the extension's options page. Second, we set the extension's content rating to "mature" on the Chrome Web Store to remind users that they could see explicit content in others' YouTube recommendations. Third, we explicitly asked users to not share identifiable information in their profile fields to reduce the risk of an information leak (see Figure 3). In case users were uncomfortable sharing any demographic details with others, the extension defaulted to not sharing any information; users manually chose what information to share with others. Similarly, we gave users the option to remove recommended videos from the collected set if they felt uncomfortable sharing them with others.

6.5   Limitations

Our study is not without limitations. First, we had a limited number of participants from the U.S. with a limited demographic distribution. Therefore, we cannot account for scenarios in which users live in different nations and speak different languages. Additionally, while many users signed up for the study, only a select few eventually participated. Therefore, some self-selection bias exists. These users could simply be those who are the most open to exchanging recommendations. Furthermore, we recruited users who use the Chrome browser on a desktop or laptop computer (required for the OtherTube extension to work) and who use YouTube regularly. These criteria likely excluded users who do not use YouTube regularly or who use it on a different platform (e.g., mobile phones). Second, our implementation of OtherTube also introduced some limits to the study. For example, some participants hesitated to watch content because preview clips were unavailable. One interview participant mentioned that they had trouble going back to a profile they had forgotten to pin. This issue could have impacted their overall interaction and engagement with OtherTube. Third, the first few participants from the first batch saw fewer profiles compared to other users; for comparison, the last person in the study had a maximum of 40 profiles to browse. Therefore, users' responses to the daily survey were affected by the limited sample. In addition to these limitations, in the future, studies can be devised

to investigate the measurable impact of systems like OtherTube on users' consumption. For example, a within-subjects study could help us quantify the difference between users' interactions with OtherTube and YouTube recommendations. Future deployment with large samples could potentially resolve these issues.

7 CONCLUSION

In this work, we investigated the exchange of social recommendations with strangers as a tool to promote content discovery and reflection on social media sites like YouTube. Our investigation revealed factors that affect users' interaction and engagement with such a system. Our work has implications for future exploration into exchanging personalized recommendations with strangers.

ACKNOWLEDGMENTS

REFERENCES

[1] Gediminas Adomavicius and YoungOk Kwon. 2009. Toward more diverse recommendations: Item re-ranking methods for recommender systems. In *Workshop on Information Technologies and Systems*. Citeseer.

[2] Shahmir H Ali, Joshua Foreman, Ariadna Capasso, Abbey M Jones, Yesim Tozan, and Ralph J DiClemente. 2020. Social media as a recruitment platform for a nationwide online survey of COVID-19 knowledge, beliefs, and practices in the United States: methodology and feasibility analysis. *BMC medical research methodology* 20 (2020), 1–11.

[3] Paul André, M. C. Schraefel, Alan Dix, and Ryen W. White. 2011. Expressing Well-Being Online: Towards Self-Reflection and Social Awareness. In *Proceedings of the 2011 iConference* (Seattle, Washington, USA) *(iConference '11)*. Association for Computing Machinery, New York, NY, USA, 114–121. https://doi.org/10.1145/1940761.1940777

[4] Mitja D Back, Juliane M Stopfer, Simine Vazire, Sam Gaddis, Stefan C Schmukle, Boris Egloff, and Samuel D Gosling. 2010. Facebook profiles reflect actual personality, not self-idealization. *Psychological science* 21, 3 (2010), 372–374.

[5] Jae-eul Bae, Youn-kyung Lim, Jin-bae Bang, and Myung-suk Kim. 2014. Ripening Room: Designing Social Media for Self-Reflection in Self-Expression. In *Proceedings of the 2014 Conference on Designing Interactive Systems* (Vancouver, BC, Canada) *(DIS '14)*. Association for Computing Machinery, New York, NY, USA, 103–112. https://doi.org/10.1145/2598510.2598567

[6] Elizabeth Bales and William Griswold. 2011. Interpersonal Informatics: Making Social Influence Visible. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI EA '11)*. Association for Computing Machinery, New York, NY, USA, 2227–2232. https://doi.org/10.1145/1979742.1979924

[7] Eric P S Baumer. 2015. Reflective informatics: conceptual dimensions for designing technologies of reflection. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, Association for Computing Machinery, New York, NY, USA, 585–594.

[8] Alessandro Bessi, Fabiana Zollo, Michela Del Vicario, Michelangelo Puliga, Antonio Scala, Guido Caldarelli, Brian Uzzi, and Walter Quattrociocchi. 2016. Users polarization on Facebook and Youtube. *PloS one* 11, 8 (2016), e0159641.

[9] Md Momen Bhuiyan, Michael Horning, Sang Won Lee, and Tanushree Mitra. 2021. NudgeCred: Supporting News Credibility Assessment on Social Media Through Nudges. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–30.

[10] Md Momen Bhuiyan, Kexin Zhang, Kelsey Vick, Michael A Horning, and Tanushree Mitra. 2018. FeedReflect: A Tool for Nudging Users to Assess News Credibility on Twitter. In *Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 205–208.

[11] Taylor C Boas, Dino P Christenson, and David M Glick. 2020. Recruiting large online samples in the United States and India: Facebook, mechanical turk, and qualtrics. *Political Science Research and Methods* 8, 2 (2020), 232–250.

[12] Lauren Valentino Bryant. 2020. The YouTube algorithm and the Alt-Right filter bubble. *Open Information Science* 4, 1 (2020), 85–90.

[13] Laura Burbach, Patrick Halbach, Martina Ziefle, and André Caleroﾠ Valdez. 2019. Bubble Trouble: Strategies Against Filter Bubbles in Online Social Networks. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 11582 LNCS (2019), 441–456. https://doi.org/10.1007/978-3-030-22219-2_33

[14] Pablo Castells, Neil J. Hurley, and Saul Vargas. 2015. *Novelty and Diversity in Recommender Systems*. Springer US, Boston, MA, 881–918. https://doi.org/10.1007/978-1-4899-7637-6_26

[15] Ying Chen, Yilu Zhou, Sencun Zhu, and Heng Xu. 2012. Detecting Offensive Language in Social Media to Protect Adolescent Online Safety. In *2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Conference on Social Computing*. IEEE, 71–80. https://doi.org/10.1109/SocialCom-PASSAT.2012.55

[16] Sheldon Ed Cohen and SI Syme. 1985. *Social support and health.* Academic Press.

[17] Sunny Consolvo, Katherine Everitt, Ian Smith, and James A Landay. 2006. Design requirements for technologies that encourage physical activity. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 457–466.

[18] Sunny Consolvo, David W. McDonald, and James A. Landay. 2009. *Theory-Driven Design Strategies for Technologies That Support Behavior Change in Everyday Life.* Association for Computing Machinery, New York, NY, USA, 405–414. https://doi.org/10.1145/1518701.1518766

[19] Paul Covington, Jay Adams, and Emre Sargin. 2016. Deep Neural Networks for YouTube Recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems* (Boston, Massachusetts, USA) *(RecSys '16)*. Association for Computing Machinery, New York, NY, USA, 191–198. https://doi.org/10.1145/2959100.2959190

[20] James Davidson, Benjamin Liebald, Junning Liu, Palash Nandy, and Taylor Van Vleet. 2010. The YouTube video recommendation system. *RecSys'10 - Proceedings of the 4th ACM Conference on Recommender Systems* (2010), 293–296. https://doi.org/10.1145/1864708.1864770

[21] Munmun De Choudhury, Michael Gamon, Aaron Hoff, and Asta Roseway. 2013. "Moon Phrases": A social media faciliated tool for emotional reflection and wellness. In *2013 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops*. IEEE, European Alliance for Innovation, 41–44.

[22] Alan Dix. 1994. Computer supported cooperative work: a framework. In *Design issues in CSCW*. Springer, 9–26.

[23] Abraham Doris-Down, Husayn Versee, and Eric Gilbert. 2013. Political blend: an application designed to bring people together based on political differences. In *Proceedings of the 6th International Conference on Communities and Technologies*. Association for Computing Machinery, New York, NY, USA, 120–130.

[24] Daniel A Epstein, An Ping, James Fogarty, and Sean A Munson. 2015. A lived informatics model of personal informatics. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. Association for Computing Machinery, New York, NY, USA, 731–742.

[25] Motahhare Eslami, Amirhossein Aleyasen, Karrie Karahalios, Kevin Hamilton, and Christian Sandvig. 2015. Feedvis: A path for exploring news feed curation algorithms. In *Proceedings of the 18th acm conference companion on computer supported cooperative work & social computing*. Association for Computing Machinery, New York, NY, USA, 65–68.

[26] Motahhare Eslami, Aimee Rickman, Kristen Vaccaro, Amirhossein Aleyasen, Andy Vuong, Karrie Karahalios, Kevin Hamilton, and Christian Sandvig. 2015. "I always assumed that I wasn't really that close to [her]" Reasoning about Invisible Algorithms in News Feeds. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 153–162.

[27] Haiyan Fan and Marshall Scott Poole. 2006. What is personalization? Perspectives on the design and implementation of personalization in information systems. *Journal of Organizational Computing and Electronic Commerce* 16, 3-4 (2006), 179–202.

[28] Leon Festinger. 1954. A theory of social comparison processes. *Human relations* 7, 2 (1954), 117–140.

[29] Clayton Feustel, Shyamak Aggarwal, Bongshin Lee, and Lauren Wilcox. 2018. People like me: Designing for reflection on aggregate cohort data in personal informatics systems. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (2018), 1–21.

[30] Feng Gao. 2012. Design for reflection on health behavior change. In *Proceedings of the 2012 ACM international conference on Intelligent User Interfaces*. Association for Computing Machinery, New York, NY, USA, 379–382.

[31] Andrew Garbett, David Chatting, Gerard Wilkinson, Clement Lee, and Ahmed Kharrufa. 2018. *ThinkActive: Designing for Pseudonymous Activity Tracking in the Classroom.* Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173581

[32] Brittany Gentile, Jean M Twenge, Elise C Freeman, and W Keith Campbell. 2012. The effect of social networking websites on positive self-views: An experimental investigation. *Computers in human behavior* 28, 5 (2012), 1929–1933.

[33] Dave Gershgorn. 2019. How Spotify's Algorithm Knows Exactly What You Want to Listen To | by Dave Gershgorn | OneZero. https://onezero.medium.com/how-

spotifys-algorithm-knows-exactly-what-you-want-to-listen-to-4b6991462c5c. (Accessed on 09/09/2021).

[34] Nabeel Gillani, Ann Yuan, Martin Saveski, Soroush Vosoughi, and Deb Roy. 2018. Me, my echo chamber, and i: Introspection on social media polarization. *The Web Conference 2018 - Proceedings of the World Wide Web Conference, WWW 2018* (2018), 823–831. https://doi.org/10.1145/3178876.3186130 arXiv:1803.01731

[35] Google. [n.d.]. Trending on YouTube - YouTube Help. https://support.google.com/youtube/answer/7239739?hl=en. (Accessed on 09/09/2021).

[36] Sten Govaerts, Katrien Verbert, Erik Duval, and Abelardo Pardo. 2012. The student activity meter for awareness and self-reflection. In *CHI'12 Extended Abstracts on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 869–884.

[37] Anthony M. Grant, John Franklin, and Peter Langford. 2002. The self-reflection and insight scale: A new measure of private self-consciousness. *Social Behavior and Personality* 30, 8 (2002), 821–836. https://doi.org/10.2224/sbp.2002.30.8.821

[38] Catherine Grevet and Eric Gilbert. 2015. Piggyback prototyping: Using existing, large-scale social computing systems to prototype new ones. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 4047–4056.

[39] Andrea Grimes, Brian M Landry, and Rebecca E Grinter. 2010. Characteristics of shared health reflections in a local community. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*. Association for Computing Machinery, New York, NY, USA, 435–444.

[40] Asela Gunawardana and Guy Shani. 2015. *Evaluating Recommender Systems*. Springer US, Boston, MA, 265–308. https://doi.org/10.1007/978-1-4899-7637-6_8

[41] Kirsi Halttu and Harri Oinas-kukkonen. 2017. Human – Computer Interaction Persuading to Reflect : Role of Reflection and Insight in Persuasive Systems Design for Physical Health Persuading to Reflect : Role of Reflection and Insight in Persuasive Systems Design for Physical Health. *Human–Computer Interaction* 32, 5-6 (2017), 381–412. https://doi.org/10.1080/07370024.2017.1283227

[42] Donald Hedeker, Stephen HC du Toit, Hakan Demirtas, and Robert D Gibbons. 2018. A note on marginalization of regression parameters from mixed models of binary outcomes. *Biometrics* 74, 1 (2018), 354–361.

[43] Natali Helberger, Kari Karppinen, and Lucia D'acunto. 2018. Exposure diversity as a design principle for recommender systems. *Information, Communication & Society* 21, 2 (2018), 191–207.

[44] Yoshinori Hijikata, Takuya Shimizu, and Shogo Nishida. 2009. Discovery-oriented collaborative filtering for improving user satisfaction. In *Proceedings of the 14th international conference on Intelligent user interfaces*. Association for Computing Machinery, New York, NY, USA, 67–76.

[45] Travonia B Hughes, Vijay R Varma, Corinne Pettigrew, and Marilyn S Albert. 2017. African Americans and clinical research: evidence concerning barriers and facilitators to participation and recruitment recommendations. *The Gerontologist* 57, 2 (2017), 348–358.

[46] Andrew Johnston, Shigeki Amitani, and Ernest Edmonds. 2005. Amplifying reflective thinking in musical performance. In *Proceedings of the 5th conference on Creativity & cognition*. Association for Computing Machinery, New York, NY, USA, 166–175.

[47] Keen. 2021. Keen | Expand your interests. https://staykeen.com/about. (Accessed on 09/07/2021).

[48] Joel Kiskola, Thomas Olsson, Heli Väätäjä, Aleksi H. Syrjämäki, Anna Rantasila, Poika Isokoski, Mirja Ilves, and Veikko Surakka. 2021. *Applying Critical Voice in Design of User Interfaces for Supporting Self-Reflection and Emotion Regulation in Online News Commenting*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3411764.3445783

[49] Kalliopi Kontiza, Olga Loboda, Louis Deladiennee, Sylvain Castagnos, and Yannick Naudet. 2018. A museum app to trigger users' reflection. In *International Workshop on Mobile Access to Cultural Heritage (MobileCH2018)*. Barcelona, Spain.

[50] Malcolm Koo and Harvey Skinner. 2005. Challenges of internet recruitment: a case study with disappointing results. *Journal of Medical Internet Research* 7, 1 (2005), e6.

[51] Elizabeth V Korinek, Sayali S Phatak, Cesar A Martin, Mohammad T Freigoun, Daniel E Rivera, Marc A Adams, Pedja Klasnja, Matthew P Buman, and Eric B Hekler. 2018. Adaptive step goals and rewards: a longitudinal growth model of daily steps for a smartphone-based walking intervention. *Journal of behavioral medicine* 41, 1 (2018), 74–86.

[52] MIT Media Lab. 2017. Overview ‹ FlipFeed — MIT Media Lab. https://www.media.mit.edu/projects/flipfeed/overview/. (Accessed on 08/31/2021).

[53] KK Lamberty and Janet L Kolodner. 2005. Camera talk: Making the camera a partial participant. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 839–848.

[54] Neal Lathia, Stephen Hailes, Licia Capra, and Xavier Amatriain. 2010. Temporal diversity in recommender systems. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*. Association for Computing Machinery, New York, NY, USA, 210–217.

[55] Matthew L Lee and Anind K Dey. 2011. Reflecting on pills and phone use: supporting awareness of functional abilities for older adults. In *Proceedings of*

[56] Ian Li, Anind Dey, and Jodi Forlizzi. 2009. Grafitter: leveraging social media for self reflection. *XRDS: Crossroads, The ACM Magazine for Students* 16, 2 (2009), 12–13.

[57] Ian Li, Anind Dey, and Jodi Forlizzi. 2010. A stage-based model of personal informatics systems. In *Proceedings of the SIGCHI conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 557–566.

[58] Ian Li, Anind K Dey, and Jodi Forlizzi. 2011. Understanding my data, myself: supporting self-reflection with ubicomp technologies. In *Proceedings of the 13th international conference on Ubiquitous computing*. Association for Computing Machinery, New York, NY, USA, 405–414.

[59] James J Lin, Lena Mamykina, Silvia Lindtner, Gregory Delajoux, and Henry B Strub. 2006. Fish'n'Steps: Encouraging physical activity with an interactive computer game. In *International conference on ubiquitous computing*. Springer, Springer Berlin Heidelberg, Berlin, Heidelberg, 261–278.

[60] Sylvain Malacria, Joey Scarr, Andy Cockburn, Carl Gutwin, and Tovi Grossman. 2013. Skillometers: Reflective widgets that motivate and help users to improve performance. In *Proceedings of the 26th annual ACM symposium on User interface software and technology*. Association for Computing Machinery, New York, NY, USA, 321–330.

[61] Sean M McNee, John Riedl, and Joseph A Konstan. 2006. Being accurate is not enough: how accuracy metrics have hurt recommender systems. In *CHI'06 extended abstracts on Human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 1097–1101.

[62] Ania Medrek. 2018. NEWS BY ASSOCIATION: Designing a way out of the echo chamber. April (2018).

[63] Susan Michie, Michelle Richardson, Marie Johnston, Charles Abraham, Jill Francis, Wendy Hardeman, Martin P Eccles, James Cane, and Caroline E Wood. 2013. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. *Annals of behavioral medicine* 46, 1 (2013), 81–95.

[64] Ine Mols, Elise van den Hoven, and Berry Eggen. 2016. Informing Design for Reflection: An Overview of Current Everyday Practices. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction* (Gothenburg, Sweden) *(NordiCHI '16)*. Association for Computing Machinery, New York, NY, USA, Article 21, 10 pages. https://doi.org/10.1145/2971485.2971494

[65] Sean Munson, Stephanie Lee, and Paul Resnick. 2013. Encouraging reading of diverse political viewpoints with a browser widget. In *Proceedings of The International AAAI Conference on Web and Social Media*, Vol. 7.

[66] Sean A Munson, Hasan Cavusoglu, Larry Frisch, and Sidney Fels. 2013. Sociotechnical challenges and progress in using social media for health. *Journal of medical Internet research* 15, 10 (2013), e226.

[67] Philip M Napoli. 2011. Exposure diversity reconsidered. *Journal of information policy* 1 (2011), 246–259.

[68] Tien T Nguyen, Pik-Mai Hui, F Maxwell Harper, Loren Terveen, and Joseph A Konstan. 2014. Exploring the filter bubble: the effect of using recommender systems on content diversity. In *Proceedings of the 23rd international conference on World wide web*. 677–686.

[69] Jeffrey Nichols and Jeon-Hyung Kang. 2012. Asking questions of targeted strangers on social networks. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*. Association for Computing Machinery, New York, NY, USA, 999–1002.

[70] Jasmin Niess and Paweł W Woźniak. 2018. Supporting meaningful personal fitness: The tracker goal evolution model. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–12.

[71] Alexander Nussbaumer, Milos Kravcik, and Dietrich Albert. 2012. Supporting self-reflection in personal learning environments through user feedback.. In *UMAP Workshops*.

[72] Adewale Obadimu, Esther Mead, Muhammad Nihal Hussain, and Nitin Agarwal. 2019. Identifying toxicity within youtube video comment. In *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation*. Springer, 214–223.

[73] Ruchi Ookalkar, Kolli Vishal Reddy, and Eric Gilbert. 2019. Pop: Bursting news filter bubbles on twiter through diverse exposure. *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW* (2019), 18–21. https://doi.org/10.1145/3311957.3359513

[74] Ruchi Ookalkar, Kolli Vishal Reddy, and Eric Gilbert. 2019. Pop: Bursting News Filter Bubbles on Twitter Through Diverse Exposure. In *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing*. 18–22.

[75] Eli Pariser. 2011. *The filter bubble: How the new personalized web is changing what we read and how we think*. Penguin.

[76] Bernd Ploderer, Wolfgang Reitberger, Harri Oinas-Kukkonen, and Julia van Gemert-Pijnen. 2014. Social interaction and reflection for behaviour change.

[77] Danielle E Ramo and Judith J Prochaska. 2012. Broad reach and targeted recruitment using Facebook for an online survey of young adult substance use. *Journal of medical Internet research* 14, 1 (2012), e28.
[78] Paul Resnick, R Kelly Garrett, Travis Kriplean, Sean A Munson, and Natalie Jomini Stroud. 2013. Bursting your (filter) bubble: strategies for promoting diverse exposure. In *Proceedings of the 2013 conference on Computer supported cooperative work companion*. Association for Computing Machinery, New York, NY, USA, 95–100.
[79] Dimitris Rizopoulos. 2019. GLMMadaptive: generalized linear mixed models using adaptive Gaussian quadrature. *R package version 0.5–1* (2019).
[80] Corina Sas and Alan Dix. 2011. Designing for reflection on personal experience. *International Journal of Human-Computer Studies* 69, 5 (2011), 281–282.
[81] Jorrit Schaap. 2020. Bubble Trouble – Venture Out of Your Filter Bubbles. (2020), 1–14.
[82] Gwendolyn Seidman. 2013. Self-presentation and belonging on Facebook: How personality influences social media use and motivations. *Personality and individual differences* 54, 3 (2013), 402–407.
[83] Swapneel Kalpesh Sheth, Jonathan Schaffer Bell, Nipun Arora, and Gail E Kaiser. 2011. Towards diversity in recommendations using social networks. (2011).
[84] Dominik A Stecula and Mark Pickup. 2021. Social media, cognitive reflection, and conspiracy beliefs. *Frontiers in Political Science* 3 (2021), 62.
[85] Anja Thieme, Jayne Wallace, Paula Johnson, John McCarthy, Siân Lindley, Peter Wright, Patrick Olivier, and Thomas D Meyer. 2013. Design to promote mindfulness practice and sense of self for vulnerable women in secure hospital services. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 2647–2656.
[86] Tiffany Tseng and Coram Bryant. 2013. Design, reflect, explore: encouraging children's reflections with mechanix. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 619–624.
[87] Nina Valkanova, Sergi Jorda, Martin Tomitsch, and Andrew Vande Moere. 2013. Reveal-it! the impact of a social visualization projection on public awareness and discourse. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 3461–3470.
[88] Anthony J Viera, Joanne M Garrett, et al. 2005. Understanding interobserver agreement: the kappa statistic. *Fam med* 37, 5 (2005), 360–363.
[89] Yixue Wang and Siyu Yao. 2020. STUDY ON INTENTION-AWARE RECOMMENDATION OF YOUTUBE VIDEOS. (2020).
[90] Mark Wilhelm, Ajith Ramanathan, Alexander Bonomo, Sagar Jain, Ed H. Chi, and Jennifer Gillenwater. 2018. Practical Diversified Recommendations on YouTube with Determinantal Point Processes. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management* (Torino, Italy) *(CIKM '18)*. Association for Computing Machinery, New York, NY, USA, 2165–2173. https://doi.org/10.1145/3269206.3272018
[91] Yan Xu, Erika Shehan Poole, Andrew D Miller, Elsa Eiriksdottir, Dan Kestranek, Richard Catrambone, and Elizabeth D Mynatt. 2012. This is not a one-horse race: understanding player types in multiplayer pervasive health games for youth. In *Proceedings of the ACM 2012 conference on computer supported cooperative work*. Association for Computing Machinery, New York, NY, USA, 843–852.
[92] Mi Zhang and Neil Hurley. 2008. Avoiding monotony: improving the diversity of recommendation lists. In *Proceedings of the 2008 ACM conference on Recommender systems*. 123–130.
[93] Shanyang Zhao, Sherri Grasmuck, and Jason Martin. 2008. Identity construction on Facebook: Digital empowerment in anchored relationships. *Computers in human behavior* 24, 5 (2008), 1816–1836.
[94] Tao Zhou, Zoltán Kuscsik, Jian-Guo Liu, Matúš Medo, Joseph Rushton Wakeling, and Yi-Cheng Zhang. 2010. Solving the apparent diversity-accuracy dilemma of recommender systems. *Proceedings of the National Academy of Sciences* 107, 10 (2010), 4511–4515.

A DISTRIBUTION OF USERS WHO PASSED ELIGIBILITY CRITERIA AND SIGNED UP

Total 318 users signed up for our study. Figure 8 shows the demography of these users.

B NEED FOR REFLECTION AND INSIGHT QUESTIONNAIRE

These questions have been taken from Halttu et. al. [41].

B.1 Need for Self-Reflection

- I am not really interested in analyzing my behavior. (R)
- It is important for me to evaluate the things that I do.
- I am very interested in examining what I think about.
- It is important to me to try to understand what my feelings mean.
- I have a definite need to understand the way that my mind works.
- It is important for me to be able to understand how my thoughts arise.

B.2 Insight

- I usually have a very clear idea about why I've behaved in a certain way.
- My behavior often puzzles me. (R)
- Thinking about my thoughts makes me more confused. (R)
- Often I find it difficult to make sense of the way I feel about things.(R)
- I usually know why I feel the way I do.

C SEMI-STRUCTURED INTERVIEW QUESTIONS

- How do you typically use YouTube to find new content?
- How would you describe the benefits and limitations in the features provided by YouTube to find new content?
- Can you walk me through your recommendation feed? How did you use features [sharing video, setting persona, browsing OtherTube] in OtherTube?
- What were your impressions when you saw strangers' profiles and recommended videos?
- Could you describe any profile/recommended videos from the strangers during the study that stood out or were memorable to you? Why?
- How would you compare strangers' recommendations to your YouTube recommendation?
- Would you continue to use OtherTube after the study? If so, what would be the purpose/motivation? If not, why not?
- Among the features provided in OtherTube about choosing your persona, what was important for you to present about yourself to others?
- How would you describe the features offered by the tool to share/remove recommended videos from your feed?
- What concerns would you have regarding sharing your profile and recommendation?
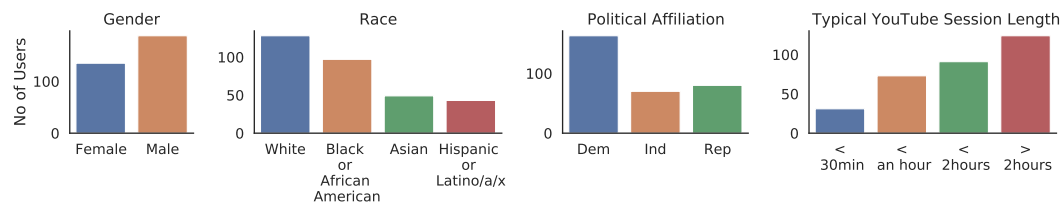- Is there anything else where the tool should be more transparent about?

Figure 8: Demography of the participants who passed screening criteria and signed up for the study.

- Other than using the plugin, was there anything that you had to do and you have not done regularly for this study?

- What did you like about the tool? What did you not like about the tool? Do you have any suggested changes on the tool for us to improve it?