# ANALYZING YOUTUBE'S PERSONALIZED RECOMMENDATIONS ON MOBILE DEVICES

*Hani Khan, Roshini Thiagarajan, Rocio Perez, Keerthana Manoharan*

## Executive Summary

Recommendations for online information is one of the most common ways to deal with large amounts of data. To keep users interested and engaged, YouTube recommends videos it thinks people would like to watch based on a variety of information it collects about user behavior. In this study, we analyze how people perceive these recommendations and how successful they find them to be. We further investigate how people go about making their decisions when it comes to selecting a recommendation and if people even prefer a personalized list over a non personalized list of videos. We conducted a user experience evaluation to gather qualitative data and then analyzed the responses through a thematic analysis to find themes in how people react to the YouTube recommender system. Our results suggest that people tend to have a sense of tolerance with bad recommendations, novelty is not always desirable, and people often struggle with a flood of irrelevant suggestions. Furthermore, we discovered that people tend to judge videos based on the title and thumbnail that is presented in addition to taking the length of the video into account. Overall, our participants were able to distinguish between personalized recommendations and non personalized recommendations.

## Introduction

The personalization of recommendations has become an essential strategy for helping users deal with information overload in systems that contain profound amounts of content. The online video sharing community known as YouTube has over one billion users that are actively watching millions of hours of video every day. YouTube also claims that "the number of users who start at the homepage, similar to how they might turn on a TV, is up more than 3X y/y"[1]. If we take a look at how the website is designed, we will see that a large portion of the home page is dedicated to recommendations. In most scenarios, users on a desktop will have a navigation pane on the left and a search bar on top. Below the search bar, users will find the main body of the page, which

[1] YouTube Statistics: https://www.youtube.com/yt/press/statistics.html

consists of a recommended videos section, followed by horizontal sections of recommended channels (Figure 1).

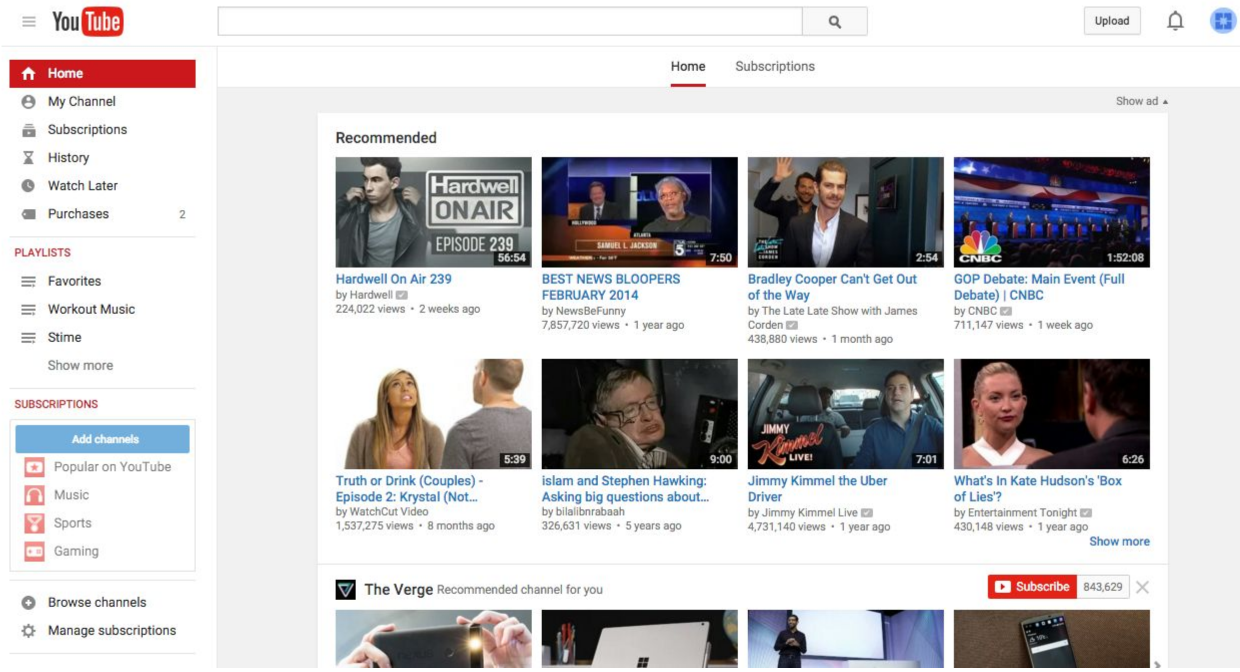
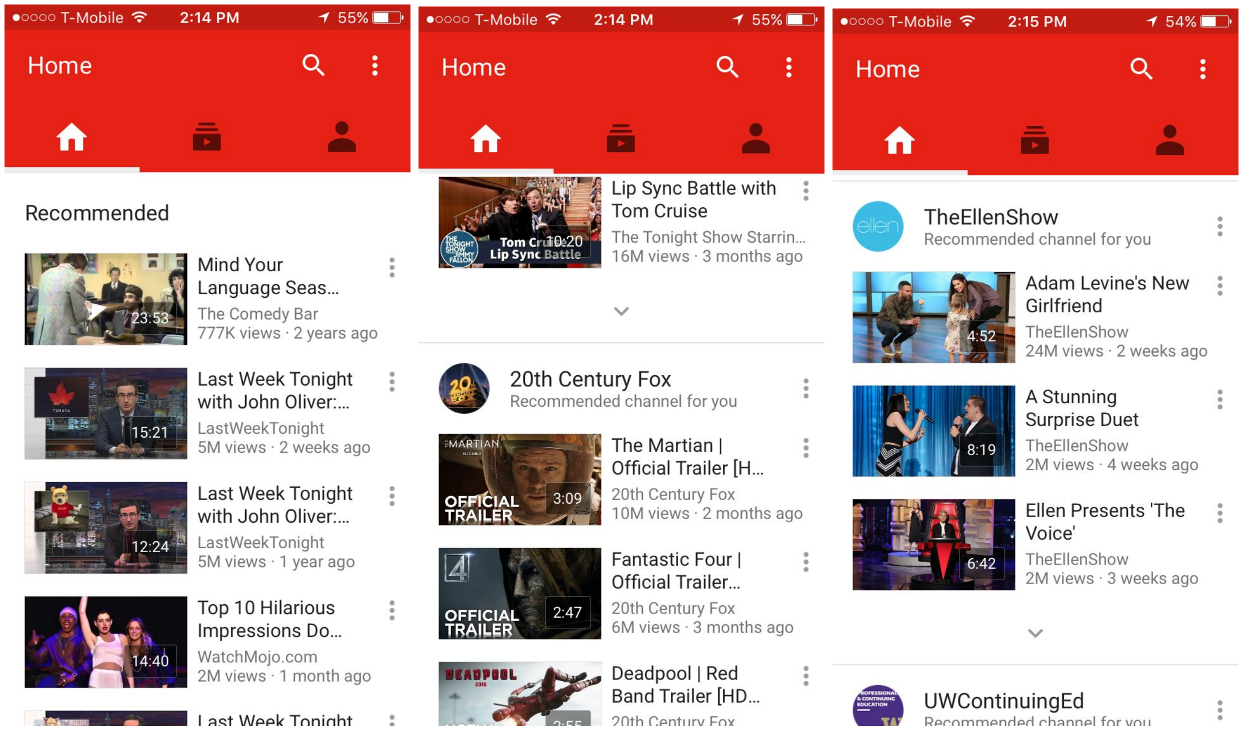



*Figure 1: YouTube Desktop Home page*

YouTube has further emphasized this focus on recommendations in their mobile application, in which the recommended videos section is the main section a user can see until they scroll down (Figure 2, 3, & 4).




*Left to right: Figure 2, 3, & 4*

The above images show how the recommended section is the first thing a user will see. Only when the user scrolls down are they able to see the channel specific recommendations. The recommended video feeds contain videos that YouTube thinks may appeal to a user. An algorithm considers several signals including a user's favorite videos, video ratings, recently watched videos, and recent searches[2]. Furthermore, a paper by Google researchers that was released in 2010 details how in addition to explicit data, implicit data is also collected such as the length at which users watch a video[3]. It's possible that this information is no longer relevant due to updates in YouTube's algorithm, however, it still goes to show the lengths to which YouTube and Google have gone to understand how their users are interacting with YouTube to further enhance the user experience. Regarding the mobile product specifically, YouTube claims that once mobile users are on YouTube, the average viewing session is more than 40 minutes and that more than half of YouTube views come from mobile devices[4]. This tells us that something about the mobile application is doing a good job at engaging users in content for long periods of time.

---

[2] Google Developers: https://developers.google.com/youtube/2.0/developers_guide_protocol_recommendations
[3] The YouTube Video Recommendation System
[4] YouTube Statistics: https://www.youtube.com/yt/press/statistics.html

Based on the information above, we can understand that there is a lot of work that has gone behind generating personalized video recommendations and that it is an integral part of how YouTube engages users. However, what is not so clear is how successful these recommendations are for people that actually click (or tap) on them. This can be tricky to track due to the wide variety of goals people may have. Therefore, for the context of this paper, we will focus on users that go to YouTube for entertainment purposes on their mobile phone. Additionally, because this study is being done in the context of a course, we had limited resources as to the time of the participants we had access to. Due to this, we put users in the context of looking for entertaining videos when they have 15 minutes to spare. We selected this context for the user study because we feel it is one that most college students can easily relate to, such as the example of watching entertaining videos while eating their lunch. Thus, the research questions that we selected are as follows: How successful do people find the personalization of Youtube's recommendations? How do people make their decisions about which recommendation to select in the context of 15 minutes to kill? Finally, we also wanted to investigate if people even prefer a personalized form of the recommendations over a non personalized form and why.

The rest of this paper will explain the methods of our study, the results we found, and a discussion of possible implications and limitations that must be considered.

## Methodology

In order to uncover how users perceive the YouTube recommender system and study how helpful the recommendations are, we conducted a user experience evaluation as suggested by Herlocker et al. (2004)[5], in which we were able to gather qualitative data. This qualitative approach helped us reach beyond the initial responses of the participant and probe for rationales. This would not have been possible in a structured quantitative analysis. In essence, this type of evaluation allowed us to inquire at length and understand why people liked or disliked specific recommendations and how recommendations affected users' decision making.

For our study, we chose to conduct a user evaluation that is a combination of the Think aloud protocol and a follow up interview. The think aloud protocol involves observing a user working on a system while encouraging them to "think-aloud", i.e to say out loud what they are thinking and wondering, at each moment while using the interface. The process involves sitting next to the users, without any special equipment,

[5] Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. ACM Transactions on Information Systems (TOIS), 22(1), 13, 45-48.

and taking notes as they talk out loud.[6] The think aloud method allowed us to record the participants' findings as and when they occur, thereby minimizing any loss of crucial data that may happen if we wait to collect information after the session. The follow-up interview allowed us to do two things: probe further into the actions of the participants during the task they were assigned and ask experiment-specific questions, the details of which the participant may not have answered while thinking out loud. Qualitative data does not require a large set of samples in order to gain important insights. Moreover, Charmaz et al[7] suggest that studies with a narrow scope may need even less samples to accomplish their goal. Therefore, we gathered data from participants until we felt that most viewpoints from the people we had access to were covered. We reached this point after interviewing 7 people total.

Participants for this study were Cornell college students selected from a graduate level Information Science class. Since our study was run on the mobile version of Youtube, it was required for them to have the YouTube application installed in their phones and an active profile (i.e. they should have used Youtube before while signed in so the recommender system could have the opportunity for personalization).

We wanted to analyze how people judge their recommendations based on a specific context. In this case, we told our users to imagine a situation in which they only have 15 minutes to spare while using YouTube on their phones. During our initial testing of YouTube, we observed that every now and then the recommended videos section appeared below a few channel recommendations. However, in this study we are only looking at instances where the recommended videos were shown at the top as it is the most common order of the screen. With this situation in mind, users were asked to find a video using the recommendations that fits the purpose and provide information about how the recommended videos affected their decisions and how successful each selected recommendation was. We observed their behaviour for 5 minutes, recorded what they said while thinking out loud and took notes about relevant and notable behaviours or opinions that arose during that time. Participants were asked to use the recommendations and select three videos and watch each for a maximum of 30 seconds to a minute.

After the interaction was over, we interviewed our users to analyze opinions and thoughts about their experience after using the system, to assess how satisfied users were with the recommender system and to address individual concerns. Questions that

[6] Thinking Aloud Protocol: http://www.nngroup.com/articles/thinking-aloud-the-1-usability-tool/

[7] Charmaz, Kathy (2006). Constructing grounded theory: A practical guide through qualitative analysis. SAGE Publications.

were asked involved a range of topics such as previous experience, accuracy and decision-making process. The questions that were used are listed below:

*Have you ever used Youtube recommendations before?*

We ask this question to find out if users have previous experience in using recommendations. Users may use Youtube to search for a certain video or they may use it to browse through their video options.

*While finding an entertaining video, how helpful were the YouTube recommendations during the task assigned?*

- *Answer choices: (not helpful - barely helpful - neutral - helpful - very helpful)*
- *What is your reasoning?* (lets the participant give an explanation for the above response.)
- *Why do you think Youtube recommended these videos?* (this is to understand if the participants perceive the recommendations as tailor-made for them or if they are presented to them based on trending videos)

*How did you go about choosing your recommendations?*

This helped us understand what aspects of the recommendation itself and/or the interface would have helped make their choices among several recommendations made by YouTube.

*What about the recommendations made you feel it was (not) helpful?*

With this question we hoped to understand what the user found helpful while making decisions about what recommendation to select.

*What information would you like to see in a recommendation?*

We asked this question to know what other data they would find helpful or interesting when it comes to choosing a recommendation. By phrasing the question in a generic, open-ended way, we hoped to gather a diverse set of responses relating to what aspects the current YouTube recommendation listing could possibly be lacking, which users would find helpful. While some may have commented about the interface, others could comment on aspects such as the diversity of recommendations.

*Now look at this set of recommendations and tell me how many recommendations do you find helpful. Tell me what you think about the two lists of recommendations.*

Here we wanted to compare the participant's personal recommendations to another user's recommendations to assess if they prefered personalized form of recommendations over a non-personalized form. The participant was expected to look at the recommendations on the researchers YouTube account and comment on how helpful they found those videos for the task at hand. Through this we hoped to understand if people are able to notice a change in personalization of recommendations and if that affects their preference.

### Approach

To analyze all the data that was collected from the think aloud observations and interviews we needed an approach that was both flexible for the context of this course and easy to learn and apply with a small group of participants. Therefore, we adopted a thematic analysis approach as this would allow us to analyze our interview responses and observations and extract meaningful information to help answer our research questions.

Braun and Clarke (2006)[8] have outlined a series of phases which researchers should go through when conducting a thematic analysis. In the first phase we read through the notes from the interviews that were conducted and noted down initial ideas that seemed to repeat in our notes. We then began generating codes regarding the topics, quotes, and observations that came up frequently or were relevant to the research questions. Once the codes were all generated, our next step was to collate the data for potential themes. We then reviewed and refined the potential list of themes. Our final step was to define and name our themes so we could extract findings related to our research questions.

## Results

To understand how people perceive the personalization of recommendations in relation to our research questions, we examined the notes that were collected from the think aloud observations and interview responses and applied the thematic analysis process to them. We have divided the presentation of our results based on the research questions that were listed in the beginning. For each research question, we list themes that were found from the analysis of the qualitative data.

[8] Braun, V. and Clarke, V. (2006) Using thematic analysis in psychology. Qualitative Research in Psychology, 3 (2). pp. 77-101. ISSN 1478-0887

**RQ1: How successful do people find the personalization of Youtube's recommendations?**

*Sense of Tolerance*

During the course of the think aloud sessions, we observed users looking past several video recommendations and telling us reasons why they did or did not like the suggestions. When asked about the individual negative recommendations, people gave us several comments such as the following:

> *Participant: "I hate it when people upload their own versions of videos and add music and stuff in the background. I just want to watch the original versions."*

In this scenario, a participant was recommended a video whose topic was relevant to his interests, however, it was a video clip from a TV Show that was edited by a YouTube content uploader. The participant expressed dissatisfaction at having this in the recommended videos section.

However, when participants were asked at the end of the interview what their overall thoughts were on YouTube's recommendations, the general sentiment was that even though there were a few recommendations that they did not like, the fact that they were able to find a successful video recommendation for the task made them feel positive towards the recommender system overall. On many occasions, participants explained that they understood why certain videos were there and were happy that it was suggested to them.

> *Participant: "I find the recommendations helpful because they, like, know what I like to watch. I don't like some of them but I can see why it's being recommended. I think I watched one of those videos briefly a while ago."*

The above example shows how a user has a very positive impression of the recommendations of YouTube even though they recognize some specific suggestions they did not like. It is specifically interesting to note the comment about how the user can "...see why it's being recommended...". This shows that users showed a sense of tolerance in many situations because they felt even the recommendations they did not like had some basis for being there.

*Novelty Is Not Always Desirable*

The concept of 'Novelty' is generally known to be an important factor in a recommender system (Herlocker et al., 2004)[9]. It makes sense in systems where a user's goal is to consume new information. However, we found that in the context of interacting with YouTube's recommendations for the purpose of entertainment, people did not always care about having novel suggestions. On several occasions, participants verbally noted that they had already seen a video that appeared in their recommendations list. However, when asked what they thought about it, they expressed satisfaction as they would be willing to rewatch the video in the current context.

*Participant: "oh I love this video! its hilarious!"*
*Researcher: "So you have seen that video before?"*
*Participant: "yes, I have."*
*Researcher: "How do you feel about it appearing in your recommendations?"*
*Participant:"I think it's awesome - I actually want to watch this right now!"*

Another participant explained how one of the videos in their recommended section was one that he repeatedly watches.

*Participant: "I watch this video a lot so I like it. I already know that this is the original video."*

The participant was happy that the video in the previous example of a famous clip from an Anime show was recommended for him.

*Flooded Recommendations*

Participants expressed how often YouTube will flood their recommendations with videos related to a video that they watched for a short amount of time and disliked.

*Participant: "There are so many videos here about hair care. I don't really care to see all those."*

In this situation the first four recommendations for the user were about hair care. In the following conversation, she mentioned that she had previously searched for a video about hair but didn't spend much time on it. She was unhappy to have these

[9] Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. ACM Transactions on Information Systems (TOIS), 22(1), 5-53.

videos suggested to her. After experiencing situations like this users also remarked on having more 'variety' in the recommendations. Participants expressed that while they were happy to see some videos related to their interests, at the same time they wish there was more diversity in the suggestions.

> *Participant: "I sometimes wish it would give me more variety in the videos here...like it's showing me all these videos from the same singer."*

This is another example where the participant's recommendations were flooded with videos that they either didn't care about or were too similar to the other videos. Participants expressed a desire to see videos that are not directly related to their previous watches but still might be of interest to them. One participant even referred to how they would like to see a section for "people who watched this, also watched...", just like the e-commerce site Amazon.

**RQ2: How people make their decisions about which recommendation to select in the context of 15 minutes to kill?**

*Book Covers Really Do Matter*

When we asked participants about which display elements were crucial in their decisions, we discovered that the thumbnail and the title were the most common elements that users considered while selecting a recommendation. Moreover, these two elements are some of the most salient visual details about each video, so we assume that the design of the interface can also influence how users pick a video, not only the quality or personal preferences.

> *Participant: "I usually look at the picture and title and instantly know if it's something I want to watch."*

The above comment was noted after the participant was asked "How do you go about choosing a recommendation?" The user went on to explain that they use the image to tell if the video is an original or if it is a fake version. Thus, users looking for original content (not mashups using different sources) can easily analyze if the video is what they want to see just by looking at the thumbnail and title. With this idea in mind and as stated by Herlocker et al, an aspect that recommender systems should consider is how the presentation of supporting information" may affect the ability of the user to satisfy their necessities and therefore, how successful these recommendations are.[10]

It is also important to note that some users made use of the visual elements to decide if the content of the videos suited their preferences. For example, some users stated that they went for a particular video because they identified the subject in the video, or even because the uploader was one they were familiar with.

> *Researcher: "So how did you go about making a decision to select a recommendation?"*
> *Participant: "Oh usually something short and uploaded by someone I trust."*

Therefore, we can conclude that in addition to the Title and Thumbnail, another strong factor involved while users select a video is the person who is the subject of the video or the content uploader.

*Length As Context Constraint*

We observed that due to the task given to the participants, people were mostly choosing videos that had a duration for less than 15 minutes. When we asked them about their decisions, they mentioned that since the task was to kill only 15 minutes, they prefer to see short videos rather than longer videos that may last more than the assigned time. This means that the context of the task has an important weight in decision making. In this case, the time constraint influenced user selections in the way they navigated through their recommendations.

> *Participant: "Oh this is a video that I really like but it's pretty long so I don't think I'll choose it now."*

The above participant made the decision to not select a recommendation simply because it was not a video he would watch in the context of 15 minutes to spare. Additionally, an intensified effect of this issue can be observed in some users. Some participants wanted to maximize this 15-minute lapse, so they tried to optimize their selections to be able to see as many videos as possible in the time frame.

---

[10] Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. ACM Transactions on Information Systems (TOIS), 22(1), 47.

Therefore, we can conclude that even though the content of the video is quite important for users, the amount of time they have can influence the satisfaction of a user's recommendations. We couldn't find any evidence to support the assumption that the length of previous watched videos are taken into consideration in Youtube recommender system, but an improvement could be applying these patterns or behaviours to provide better recommendations.

**RQ3: Do people prefer a personalized form of the recommendations over a non personalized form?**

The last part of the interview was to show the participant a list of recommendations that were not personalized for them and ask them to comment on it. For most participants, the response that we received was that they preferred their own list of recommendations because they felt it was much more relevant to their interests. Participants were immediately able to express their dissatisfaction in the non personalized list that was shown to them.

> *Participant: "The list on my phone is much better. It's tailored to my interests. There are so many more funny things here. The other list is fine, i'm just not interested in any of them."*

In some scenarios, participants commented that the non-personalized recommendations had a few videos that looked interesting to them. However, we believe this to be a consequence of interviewing students that are in the same social circle. We came across coincidences where both the participant and the researcher had some similar interests. We believe that with a more diverse set of participants, the result of a preference for the personalized list of recommendations would be even stronger based on the patterns we observed from our thematic analysis.

## Discussions & Limitations

There are several limitations that should be considered when interpreting our findings. First, the 15-minute time constraint of the given task could have influenced some of our participants to choose videos that are shorter than what they would have chosen if there were no limits on the time spent with the system. In many cases, our participants told us they weren't selecting a particular video because we gave them a time constraint. Although this shows that users, in general, are mindful of the length of the video, our scenario might not be the natural way in which they use the system to find entertaining videos.

Second, we noticed that many participants were quite understanding when the system flooded their recommended section with recommendations that were very similar. Having had some exposure into how the recommender systems work, our participants knew that Youtube is continuously tracking their usage pattern, is learning from these patterns and is trying to serve those recommendations it thinks would appeal to the user. This knowledge might have led them to have a more lenient view towards Youtube recommendations.

Thirdly, the participants of our study, during the last section of the interview were able to find helpful and interesting recommendations from the list that was personalized for the researcher and not for themselves. From our follow-up questions, we realized that these coincidences could have been due to the participants and the researchers being drawn from similar age groups and social circles. The research, if conducted with a more diverse participant demography, may yield different results.

Finally, for this study, we specifically asked the participants to only choose videos from the recommended section on the homescreen. Many participants indicated that they often use this section to select a particular recommended video. However, once they are on a video screen, they use the related recommendations on the right side of the screen to video-hop. Since, we were only evaluating the homescreen recommendations, we did not consider the recommendations listed on the video pages themselves. This may not be the usual way in which users navigate through the recommendations on a regular basis and thus, could have influenced our findings.

## Conclusion

Keeping the users engaged with the system is a major priority for YouTube. Recommendations play a huge role in helping YouTube keep up with its priority. This was apparent from the thematic analysis of our research study, where in our participants in general had a positive image about the YouTube recommendation system. First, our qualitative study suggested that users are tolerant towards a few bad recommendations, are not always looking for novelty and are unhappy about flooded recommendations. Second, our study indicated that users judge the recommendations based on prominent visual cues including the title, thumbnails and the length of the video. Finally, our study suggested that users prefer personalized recommendations over non-personalized recommendations when given the option to compare the two.