YouTube Needs: Understanding User's Motivations to Watch Videos on Mobile Devices

Rodrigo de Oliveira
YouTube
San Bruno, CA, USA
oliveirar@google.com

Christopher Pentoney
YouTube
San Bruno, CA, USA
cpentoney@google.com

Mika Pritchard-Berman
YouTube
San Bruno, CA, USA
mikapr@google.com

ABSTRACT

Watching online videos on mobile devices has been massively present in people's daily activities. However, different users can watch the same video for different purposes, and hence develop different expectations for their experience. Understanding people's motivations for watching videos on mobile can help address this problem by giving designers the information needed to craft the whole watching journey better adapting to user's expectations. To obtain this understanding, a comprehensive framework of viewer motivations is necessary. We present research that provides several contributions to understanding mobile video watchers: a thorough framework of user motivations to watch videos on mobile devices, a detailed procedure for collecting and categorizing these motivations, a set of challenges that viewers experience to address each motivation, insights on usage of mobile and non-mobile devices, and design recommendations for video sharing systems.

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

Author Keywords

YouTube; video; motivations; experience sampling method.

INTRODUCTION

Online video consumption is a key part of people's digital daily interactions. Only on YouTube, users spend more than a billion hours watching videos each day¹. It is estimated that by 2021, 80% of all Internet traffic of 3 zettabytes should come from video². This huge video consumption has driven business not just for professionals, but for users around the world who create and share video individually aiming to one day become as successful as YouTube celebrities³.

¹Available at: <https://goo.gl/2RcsAH>, Feb 2017.

²Available at: <https://goo.gl/QLMgK9>, Jun 2017.

³Available at: <https://goo.gl/tv449W>, Dec 2016.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

MobileHCI '18 September 3–6, 2018, Barcelona, Spain

© 2018 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-5898-9/18/09.

DOI: <https://doi.org/10.1145/3229434.3229448>

Much time, effort, and money goes into the design of online video sharing systems, with several new products and services being built in the past few years⁴. However, these systems and key stakeholders lack understanding of all key user motivations for watching videos. If designers of these systems knew the different reasons why people watch videos, then they could design for each motivation and related use case. Similarly, if content creators had a detailed picture of how each of these motivations contribute to their audience's interest in their content, they could craft content optimizing for video consumption.

Previous research typically focused on describing specific content areas [9, 13] and experiences [4, 10, 11, 12, 14, 15, 16, 17] of viewer motivations. Important frameworks for understanding viewer motivations were developed [3, 7, 8, 13]. However, these frameworks fall short in at least one of three aspects. First, they often focused on specific contexts that prevent a broader generalization of their findings for mobile video consumption. Second, their data were usually collected with a methodology that demands high recall from users, such as deploying a survey that can reveal most frequent viewer motivations, but hides less frequent—yet important—ones. And third, they do not combine qualitative user perception data with quantitative logs of mobile user behavior data to provide an in-depth understanding of each viewer motivation. Thus, a complete model of viewer needs in the mobile space had not been established.

To develop a more comprehensive framework, we conducted a large scale study with 432 participants that combined data from user's state of mind, context, behavior, perceptions, and experience. This provided: (1) a framework of motivations to watch videos on mobile, (2) a detailed procedure for categorizing these motivations, (3) a set of key challenges that today's users face when trying to address their video watching motivations while on the go, (4) insights on how mobile and non-mobile devices are used to cope with these motivations, and design recommendations.

The next section presents related work. We then describe our methodology in detail to support future studies. Results and discussion follow, presenting the framework of mobile video watching motivations, challenges to address them, and usage of mobile and non-mobile devices for that. Finally, we present

⁴Available at: <https://goo.gl/5ojA1f>, May 2016.

key implications for the design of video sharing systems that were derived from this work.

RELATED WORK

Several studies used *frameworks* to describe video-watching motivations. In 2006, Lagger et al [7] conducted interviews and surveys to describe viewer motivations. This framework included intentions of entertainment, boredom, music, information, product reviews, news, and learning. Of these, people most often reported entertainment and music. The study featured a taxonomy of entertainment and informational motivations, but did not include motivations to connect with other people as part of the framework. Bondad-Brown et al. [3] included companionship, and compared motivations for online video watching to those of traditional television. Motivations for information were found to be relatively strong for online watching compared to television. Furthermore, emotional motivations were consistently important across studies [3, 7, 8, 13]. Using 16 basic human motivations (e.g., power, knowledge, companionship, etc.) from sensitivity theory, Reiss & Wiltz [13] explained why people watch reality TV. Individuals who indicated stronger motivations for social status and vengeance were found to be more interested in reality TV shows. A 2017 study by McNally & Harrington [8] explored a set of 13 motivations for Millennials and teens watching video. The study employed a diary study method combined with interviews to describe different scenarios in which younger people were watching videos. All of these studies [3, 7, 8, 13] contributed theoretical work for several video-watching motivations, but could not provide an in-depth framework of motivations for a couple of reasons. For example, using methods like deploying one survey and semi-structured interviews demand high recall from users and tend to reveal only the most frequent motivations. Experience sampling methods that probe users in closer temporal proximity to their latest video watch experiences are key to provide a more comprehensive framework on motivations to watch videos. Also, previous studies have not combined qualitative data on user's perceptions of their motivations with quantitative behavioral data to get more detailed and actionable frameworks for mobile video watching motivations.

Other research focused on the *context-specific* aspects of video watching. Research provided strong evidence that viewers were motivated to view videos that were likely to have a desired emotional impact [9, 12]. Pirouz et al. [12] conducted a study on 622 YouTube videos to measure the relationship between a video's content format and viewer engagement (likes, comments, and views). Emotionally surprising videos had the highest engagement measures [12], which was consistent with other studies that described the importance of emotional motivators [7, 13]. Recently, Myrick [9] examined the reasons people watch cat videos on the Internet. One intention people had when watching these videos was to improve their mood [9]. Though informative for emotional motivations, these studies did not provide data on a wide range of motivations, especially those of key importance to mobile video watch experiences.

Motivations to obtain *information* from a video were explored in two of the previous video watching framework studies [3,

7]. In 2012, Bondad-brown et al. [3] found that informational motivations may be more common online than on TV. The implication was that people used online video as an instrument to obtain something specific, in contrast to TV, which was a more passive activity. However, user intent when looking for information may be more nuanced than that. To what degree a user knows what they are looking to watch may depend on the precise reason the viewer is seeking information. For example, Lagger et al. [7] summarized informational motivations as belonging to two separate groups: information search and education. Conceptually, these are related to user intent. Information search is likely to have clear user intent, but when users are looking to be educated, they may not yet have enough information to conduct an exact search. Nonetheless, users could engage in information search for a number of reasons, such as to get product reviews or news broadcasts, (which Lagger et al. had used as survey items [7]). Overall, there is a need for expanding research on informational motivations, understanding how mobile devices support them and how clarity of intent influences user's expectations.

Contrastingly, research on motivations for *connection* was usually very context or application specific. There was a variety of applied research on watching videos in social contexts, which was primarily framed around a desire to connect with others. The prevalence of applied research on social motivations highlighted the importance of including a companionship motivation in a video watching framework. Research related to such motivations was focused on many real-world applications. Examples included studies on synchronization requirements for co-watching videos [4], guidance to video creators for making shareable videos [12], development of new methods for editing and sharing panoramic videos [10], and recommendations for easing the challenges of co-watching videos on YouTube [16]. Research conducted on social context [4, 10, 12, 16] was very focused on specific applications of the understanding of motivations to connect. Unfortunately, these studies were not intended to provide descriptions of a taxonomy of viewer motivations; they each focused on their respective aspects of user needs for companionship. Nonetheless, the breadth of these applied studies indicated the importance of the connection motivation. Research is still needed to investigate how this and other types of video watching motivations are addressed in the mobile space together with user's key challenges.

Substantial work has been conducted to categorize and understand specific user motivations for watching videos, particularly on YouTube given its massive use⁵. However, a comprehensive description with insights on user's state of mind, context, behavior, perceptions, experience, and challenges when addressing their mobile video watching motivations has not yet been developed. Given this opportunity, the following questions were the focus of this study:

1. How can mobile viewer motivations be comprehensively categorized, studied, and compared on YouTube?
2. How often do these motivations occur?
3. How do users experience them (behavior, context, etc.)?

⁵Available at: <https://goo.gl/2RcsAH>, Feb 2017.

4. What mainly changes when using non-mobile devices?
5. What challenges are faced to address these motivations?
6. What can designers do to cope with these challenges?

METHODOLOGY

We used a mixed-methods approach to identify users' motivations and challenges to watch videos.

First, we used the Experience Sampling Method (ESM), which involves asking participants to report on their experiences at random points throughout the day. This method was originally proposed in Psychology [1] and later adapted in Human-Computer Interaction [6, 5]. The main advantage of this method is that it preserves ecological validity of the measurements since users are "probed" in close temporal proximity to when a relevant event is produced. By instrumenting the random probes with mobile phone notifications to answer short surveys each day, we scaled data collection increasing statistical power of our analyses. From now on, we call them *probing surveys*.

Second, we used logs analysis by combining qualitative data from the probing surveys with their respective quantitative user behavior logs on YouTube. That helped us connect user perception of their watch experience with what they actually did on YouTube.

We then manually inspected the merged dataset, evaluated different categorization schemes, and finally categorized user's motivations for every watch experience reported in the probing surveys. This enabled us to reveal user's needs and challenges to watch videos, as well as quantify different aspects for them.

Finally, we conducted follow-up semi-structured interviews with a small subset of participants from the larger ESM study, aiming to get more qualitative depth on reports about challenges to address video watching needs reported in the probing surveys.

Next, we give details on the profile of the participants of our study, the apparatus used to collect data, and the procedure to deploy the study. Future research can take advantage of this procedure to sample and categorize user motivations.

Participants

Participants were recruited through a major recruiting agency in the US. A total of 432 participants (171 male, 39.6%) fulfilled the eligibility criteria and completed all requirements of the ESM study. Participants were eligible to participate in the study if they: (1) had either an iOS or Android phone that they could install the ESM app to receive mobile notifications for answering the surveys; (2) had or were willing to create a Google account to login to YouTube during the study in order for us to collect their logged behavior; and (3) had used YouTube in the past. Also, we recruited for an even distribution of highly-engaged and low-engaged YouTube users to collect enough data points in each group and draw comparisons between them. Highly-engaged users were defined as those who reported using YouTube about every day, reported spending at least one hour per week on YouTube, and did not

perceive themselves as casual YouTube users. Participants ranged in age between 18 and 50 years old ($\bar{x} = 30.0$, $s = 6.0$). Our sample covered all nine geographical regions in US. They came from a diverse range of backgrounds, and 80.3% were either self-employed or full-time/part-time workers. Each participant received a monetary compensation for participating in the whole study.

Apparatus

We implemented the ESM study using PACO, a free open-source mobile app that supports the design, administration, and data collection of experience sampling studies [2]. Since it was available solely for iOS and Android mobile devices, it was a requirement of the study to own one of these devices. At the time the study was conducted, 95.6% of US residents who participated in a nation-wide survey⁶ reported owning a mobile device with one of these platforms, which indicated that this requirement did not significantly constrain the sample.

Procedure

The initial pool of candidates that met the eligibility criteria filled out an online survey to provide more information about their profile. Candidates that reported being available during the 3 weeks of the study (August 2015) were then randomly selected to participate. They were contacted by phone two weeks before the study to confirm availability and clarify requirements of the study. Next we detail the different phases of the study.

Phase 1: Onboarding. This phase lasted one week. Participants were sent onboarding instructions via email. These included how to create a Google account (in case the participant did not have one), how to login to YouTube using that account in every device they had (e.g. mobile, desktop, tablet, TV), and a video tutorial detailing steps needed to download PACO on their phone, join the study in PACO, and how to participate in the study. If participants expressed difficulties completing onboarding steps, we called them and provided step-by-step guidance. No data collected in this initial phase was used in the final analysis.

Phase 2: ESM Study. This phase was the actual ESM study and lasted for 2 weeks. Every day, users received 3 notifications at random times to answer a probing survey (see Table 1 for a summary of the main questions that we asked to participants in each probing survey, along with their associated measures). The survey inquired participants about the last time they listened or watched to a video on YouTube, with follow-up questions about the experience. Participants could dismiss or postpone the survey. They could also voluntarily fill the survey at any time to report a video watch without waiting for a probe. By using this method, we collected information about participants' video watches at close proximity to when the experience happened (usually not more than a few hours after the event), thus making it more likely to capture an infrequent motivation in comparison to using single surveys [13, 3, 7, 16]. Figure 1 shows examples of the ESM notification and probing survey.

⁶<http://www.businessinsider.com/apple-iphone-market-share-in-the-us-v-android-2015-12>

Figure 1. Example of ESM notification received through PACO (left) and answers to the corresponding probing survey (right).

Phase 3: Follow-up Interviews. When the study finished, we inspected the probing survey data to spot cases when participants reported few information about their motivation to watch videos and challenges faced. We recruited 10 of these participants for a follow-up interview.

Phase 4: Logs Matching. We retrieved participants' YouTube logs from the study period and matched them with the corresponding probing surveys. Matches were based on three variables: (1) Account: matching the participant's email account from the probing survey with the one associated with their YouTube logs; (2) Date/Time: matching the period of the day that the participant reported watching videos with the logged timestamp of videos they watched; and (3) Content: finding similarities between the actual videos watched (as recorded in the YouTube logs) and what participants reported in the survey about what they were looking for. First, we wrote and ran a script that listed all of the potential matches based on account and date/time. Then, we manually inspected the potential matches, confirming the accurate ones. This process allowed us to create a robust dataset with detailed qualitative and quantitative data about people's needs, motivations and behavior when watching online videos.

Phase 5: Framework of Motivations. A group of researchers including the authors manually inspected random parts of the final dataset to identify clustering patterns for the participants' self-reported motivations to watch videos. We conducted workshops where each researcher proposed their clustering approach and the others spotted pros and cons for each one. Our key goals for the final framework were: (1) Consistency: for partitioning clusters; (2) Ease of coding: for feasibility and reliability in future studies; and (3) Usefulness: for a good balance of detail to support research in the Human-Computer Interaction community. Also, we avoided labeling motivations as content categories (e.g., politics, sports), and instead la-

beled them according to user's self-reported motivation. For example, two users might have answered "*political debate*" for what they were looking for, but gave different reasons why, such as "*to be updated on what's going on*", and "*just to kill time*". Although the content watched was similar (political), motivations were different. The final framework was agreed upon by all researchers and the dataset was split amongst them for manual coding. Reliability results were good for the first attempts (20 cases, 5 coders, Krippendorff's $\alpha = 0.75$) and even higher after coders met and clarified coding instructions (K-alpha = 0.81). The detailed classification scheme for video watching needs is presented in Table 2.

RESULTS AND DISCUSSION

Results in this section are related only to user reported cases of watching YouTube on at least one mobile device, particularly phones, tablets, and/or wearable devices, unless otherwise stated. Overall, 6,520 cases of watching YouTube on mobile devices were reported with the probing surveys, with 3,749 being from highly engaged YouTube users. Generally, patterns of the highly-engaged and more casual users were similar. Because of this, we present data for the highly-engaged users to keep results concise. When observations for highly-engaged and casual users differ significantly, we contrast data between these groups. Each case from the probing surveys was manually classified into three themes of motivations to watch videos: Entertainment, Information, and Connection. We further expanded these broader categories into 14 specific user motivations addressed on mobile devices. Table 2 shows the framework of user motivations for watching videos on YouTube.

Next we present detailed results for each of these 14 motivations under the Entertainment, Information, and Connection categories.

Motivation of Entertainment

Of reported cases of using YouTube with mobile devices, the majority (59.5% were classified as being for entertainment. The relatively high rate of entertainment motivations was expected, as these cases were quite prevalent in previous literature [3, 7, 13]. Entertainment cases included two types of specific visits to YouTube: visits where users were watching videos for a topic they felt passionately interested in (41.7% of entertainment visits) and visits where users were pursuing an emotional state (49.5% of entertainment visits, see Table 3). There was a subset (8.8% of entertainment visits) of cases that provided enough context such that they could be classified as entertainment, but not enough information to categorize more specifically. In almost every entertainment motivation, highly-engaged YouTube viewers spent significantly more time watching videos compared to lower-engaged YouTube viewers. This was expected per definition of engagement. Next we give details about each of these entertainment motivations addressed with mobile devices.

Enjoy an Interest

Enjoying an interest was amongst the most frequent motivations to watch videos, and the most frequent for entertainment

Questions and Responses	Type	Notes for analysis
When did you last watch or listen to a video on YouTube? 1. I watched YouTube last night (yesterday 8pm - today 5am); 2. I watched YouTube this morning (5am - 12pm); 3. I watched YouTube this afternoon (12pm - 5pm); 4. I watched YouTube this evening (5pm - 8pm); 5. I watched YouTube this night (8pm - 12am); 6. I haven't watched YouTube since last night 7. I haven't watched YouTube since the previous survey.	Single choice	
How clear was your intent? 1. I knew exactly what I wanted to watch; 2. I had an idea of what I wanted to watch; 3. I had nothing in mind to watch.	Single choice	We report "clear intent" the responses for option #1.
Describe what you were originally looking for (Please enter a full sentence)	Open	
Briefly describe WHY you wanted to watch that video/TV/movie/etc.	Open	
Check all devices that you used 1. Desktop or laptop; 2. Phone; 3. Tablet; 4. TV; 5. Game console; 6. Wearable (e.g. smart watch); 7. Other	Multiple choice	We report "mobile" the responses for options #2, #3, #6; "TV" for #4 and #5
How quickly did you need to fulfill this need? 1. Needed immediately; 2. Needed within a few minutes; 3. Needed within 1 hour; 4. Needed within a few hours; 5. Needed within 1 day; 6. Needed within a few days; 7. Needed within 1 week; 8. Not urgent at all	Single choice	We report "urgent motivation" the responses for options #1 and #2.
Who else was watching with you? 1. Watched alone with nobody around me; 2. Watched alone with people around me; 3. Watched together with one person; 4. Watched together with a group of people.	Single choice	
Briefly describe the main challenge (if any) to fulfill this need on YouTube (or things that could be improved).	Open	
Check all things you used to fulfill this need: 1. YouTube; 2. Search engine (e.g. Google); 3. Social sharing (e.g. Facebook, Snapchat); 4. Other video streaming (e.g. Netflix, Amazon); 5. Other website w/ video (e.g. news, blogs); 6. TV (e.g. cable, satellite); 7. Cinema or outdoor projection; 8. Recordings (e.g. DVD, VHS); 9. Other.	Multiple choice	

Table 1. Key measures collected with the probing surveys.

Category of motivation	Motivation	Ask when categorizing the motivation	Typical keywords	Example	% Occurrence within category
Entertainment	Enjoy an Interest	Are you trying to be entertained by enjoying content of a specific interest?	Enjoy, like, love	"I love to watch Walking Dead when it comes on every Sunday"	41.7%
	Relieve Boredom	Are you needing to change your mood to relieve boredom?	Kill time, waste, browse	"I was bored and needed to find a way to pass the time"	27.7%
	Feel Happy	Are you needing to change your mood to feel happy or better?	Laugh, feel better	"I was hungover and wanted to have a laugh to feel better"	8.9%
	Focus	Are you needing to change your mood to help you focus on a task or facilitate it?	Concentrate, exercise, work	"Music helps me to focus and clean the house"	6.1%
	Relax	Are you needing to change your mood to feel relaxed?	Wind-down, calm, de-stress	"I had a long day and just need to unwind"	2.8%
	Be Inspired	Are you needing to change your mood to feel pumped up?	Pump up, excite, get ready	"I was getting ready for the day and needed a pick me up"	2.2%
	Be Nostalgic	Are you needing to change your mood to go back to a feeling you had?	Oldies, past, comfortable	"I wanted to watch 80's movies to remind me of my childhood"	1.8%
Information	Learn for Self-development	Are you trying to learn something for personal education or for work/school?	Curious, learn, explore	"I'm curious about psychology and want to learn more about it"	41.8%
	Get Instructions to do Something	Are you looking for instructions on how to fix/make/do something?	DIY, fix-it, step-by-step	"I got a paper jam in my HP Envy 4500 printer, and I need to fix it"	35.8%
	Make a Decision	Are you trying to make a decision?	Trailer, review, comparison	"I'm thinking of buying the new Nexus 5 and want to watch some reviews"	15.3%
	Get the Latest	Are you looking for latest information like news or updates?	News, fresh, update	"I wanna watch the news to see what is going on"	7.1%
Connection	Proactively connect	Did someone have a need that you decided to address? Or did you create a need for someone else?	Share, help, facilitate	"I want my kids to watch more educational videos"	49.1%
	React to connection	Did someone else ask you or expect you to do something?	Watch, respond, appease	"My friend sent me a video, so I watched it"	42.2%
	Mutually connect	Did you and someone else have a need to do something?	Together, connected, two-sided	"We love to watch funny shows together at night. It's our bonding time."	7.7%

Table 2. Framework of video watching motivations with detailed classification scheme. Frequencies of observations within each category from our study per motivation are presented in the last column.

Motivation	% Occurrence	User State of Mind	
		% Clear intent	% Urgent
Enjoy an Interest	41.7%	69.6%	51.7%
Relieve Boredom	27.7%	13.9%	45.2%
Feel Happy	8.9%	46.2%	63.8%
Focus	6.1%	46.7%	80.3%
Relax	2.8%	32.3%	75.8%
Be Inspired	2.2%	70.8%	77.8%
Be Nostalgic	1.8%	57.5%	77.5%

Table 3. Description of Entertainment motivations. Occurrences are percentages of all Entertainment motivations addressed with mobile devices. A subset of 8.8% of cases was not further categorized within Entertainment due to the lack of detail in the related user reports.

(41.7%). This is a key user need that YouTube and other video platforms aim to fulfill. Participants usually had a clear idea of what they were looking for (69.6%), as in this example: *"I wanted to listen to a song by The Fray, because I love their music"*. This indicated that improved forms of search and customization of favorite content are quite relevant to address this motivation. Also, enjoying an interest seems to be related to higher expectations of quality. Quality was one of the most reported challenges when users had this motivation, e.g. *"I found the video that I wanted but it looked like it was a video of a video someone took on their home TV. I was hoping a VEVO version or a verified version"*. Since the content and entity around it is so important to the users, they care a lot about distinguishing sources. Additionally, similar to previous findings [7], ads experience requires a lot of care since enjoyment of the full content is key to the experience, e.g. *"The main annoyance was the cut-ins of the ads. I literally saw it 3 times and it cut right in the middle of a crucial part of the video"*. Participants often used YouTube alone to enjoy videos of their particular interests (65.7% of cases). Also consistent with previous findings [7], they most often watched videos of music, vlogs, and/or general entertainment. Participants were rarely using other tools along with YouTube to enjoy an interest, with the most common tool used being search engines in 14.1% of cases. This is another indication that enjoying an interest is mostly a motivation with a clear goal.

Relieve Boredom

Watching videos to relieve boredom was another frequent case of entertainment (27.7%), which is in agreement with previous work [7]. It happened mostly when participants were alone (66.8%), and participants often did not consider it a need to be urgently addressed (45.2%). Relieving boredom is the main example of unclear intent, as exemplified by this participant: *"I was looking for anything because there was nothing else to do"*. In fact, in only 13.9% of cases participants reported knowing exactly what they were looking for. This motivation puts users in a state of uncertainty where they lack direction: *"It's hard to find something when you don't know what you're looking for"*. Since intent is unclear, users seem open to explore content and they complain when results are not diverse enough: *"All suggested videos are the same type of content and not a true representation of everything I watch"*. Unclear intent is a big challenge for video search engines because search terms are

usually broad, leading to suboptimal search results that can overwhelm users, as described by this participant: *"[It's] too many choices"*. Common video genres included entertainment, vlogs, and comedy. Users often used social networks (33.7%) along with YouTube when addressing this motivation, which could indicate that users have to try several approaches to properly fulfill this need.

Feel Happy

Given its relatively high occurrence amongst entertainment motivations (8.9%), feeling happy was found to be a very desirable emotional state. Like the other emotional states, participants reported urgency to achieve it (63.8%). The high urgency may be related to the fact that participants often experienced this motivation as an escape from an undesired previous emotional state, as found in previous research [9] and as expressed in the following example from this study: *"I was looking for funny videos and music because I was in a bad mood"*. Subjectiveness of humor is a key challenge to address this motivation. Content variety also plays an important role: *"I wanted to find something new, but ended up going back to my usual"*. It was often reported to be a task performed alone (62.8%). The most common genres watched were entertainment, comedy, and vlogs. Users didn't often report using other tools along with YouTube; only 13.6% of the cases included social network use, and 9.6% included the use of a separate search engine.

Focus

Participants either watched or listened to YouTube videos to help them focus on another task in 6.1% cases of entertainment, e.g. *"I was looking for music to listen to while cooking"*. Users were typically alone (64.2%) when they were looking for a video to stay focused. Interestingly, video watches that addressed this motivation at work occurred significantly less often when mobile devices were used than when they weren't (13.1% vs. 50.8% respectively, $\chi^2 = 30.2, df = 1, p < .001$). This might indicate that users tend to leverage more their work computer than their mobile devices to watch or listen to videos in order to focus on a certain task at work. Similar to findings from Lagger et al. [7], music was the most common genre of video watched. Users rarely used other tools along with YouTube (search engines were used only 13.1% of the time). Listening to content while doing something else is a lean-back experience, and therefore participants were not interested in investing too much time to have more control over it: *"I just wanted something to keep playing without me having to do anything"*. Although not amongst the most frequently reported motivations (6.1% of entertainment), participants spent a lot of time addressing this motivation. Because of that, users tended to listen to similar music lists multiple times. Encountering new good music for focusing becomes a significant effort.

Relax

Participants reported watching videos to relax and wind-down in 2.8% of entertainment cases: *"I wanted to listen to music to help me go back to sleep"*. It was a task often performed at night (62.9%), and users almost always reported being alone (80.7%). In previous research [7], this motivation was conceptually combined with motivations of either entertainment or

boredom. However, there were unique considerations when users were specifically trying to relax: Ads experience was a sensitive matter since loud or funny ads were reported to disrupt relaxation. These results suggest that contextual ads could play an important role improving the user's experience when addressing this motivation (e.g. selecting ads with relaxing/calm audio content during later periods of the day). Also, discovering good longer content can be tedious: *"The hard thing was finding a long enough video to sleep to"*. Entertainment, music, and video blogs were the most common types of content watched. Other tools were generally not used along with YouTube.

Be Inspired

Watching videos to feel inspired occurred in 2.2% of entertainment cases. These were addressed most often during mornings (41.7%), as depicted by this participant: *"I was getting ready for the day and needed a pick me up"*. Given the broad nature of searches for inspirational videos, results often would not properly address the need, but rather overload users with options: *"There are too many videos on this topic. Would be nice to have it categorized"*. Users need help selecting amongst curated content. Also, videos and compilations were expected to have high quality in order to be inspirational: *"It was hard to find something with decent audio"*. To be inspired, users generally reported watching videos alone (64.6%). Common categories of content were entertainment, music, and video blogs. Participants sometimes used search engines (12.5%) along with YouTube when trying to be inspired.

Be Nostalgic

User motivation to be nostalgic was reported in 1.8% of cases when watching videos for entertainment. Self-report methods from other research [3, 13] may not have captured this motivation due to its rarity. Feeling nostalgic requires reliving something from the past. Even though participants usually knew exactly what they were looking for (57.5%), finding these old videos was sometimes challenging, as explained by this participant: *"The videos are a few years old so a lot of irrelevant videos came up"*. In addition, older content is often lower quality, which interferes in the user's goal to feel nostalgic: *"The main challenge is finding a good quality video that wasn't a recording on VHS"*. The motivation to be nostalgic was most often pursued alone (65.0%). Music was the most common type of content. Users reported sometimes using search engines (20.0%) along with YouTube to help feel nostalgic. This was the highest use of search engines for entertainment purposes, which confirms that video was not the primary form of nostalgic content for users.

Motivation of Information

Visits to obtain information were 30.7% of reported visits to YouTube using mobile devices. The somewhat low frequency of informational motivations was consistent with some previous work [7], but contrasted with findings from other research [3]. When watching videos for information, people usually had more clear intent (65.0%), and were often urgent (54.4%). However, these visits were less urgent overall for lower-engaged users (39.7%) than for highly-engaged users

($\chi^2 = 45.5, df = 1, p < .001$). Informational visits were further classified into four separate motivations: learning for self-development, getting instructions, making a decision, and getting the latest information about a topic (See Table 4). Next we present details on each of these information motivations addressed with mobile devices.

Motivation	% Occurrence	User State of Mind	
		% Clear intent	% Urgent
Self-development	41.8%	58.6%	47.0%
Instructions	35.8%	69.5%	62.6%
Decision	15.3%	72.2%	44.9%
Latest	7.1%	64.6%	76.8%

Table 4. Description of Information motivations. Occurrences are percentages of all Information motivations addressed with mobile devices.

Learn for Self-Development

When watching videos for information, participants most often did it for self-development (41.8%), as reported by this participant: *"I was looking for videos on CERN because I wanted to learn more about it"*. Consistent with research on videos watched in social contexts [17], watching for learning motivations was a task mostly performed alone (65.1%), covering mainly blogs and entertainment video categories. One of the main challenges reported by our participants was the inability to check the legitimacy of the source watched: *"[the challenge was] gathering true information on the candidate"*. Since topics for self-development are often broad, search results may not be very accurate: *"It was difficult to find additional videos on the exact topic I wanted"*. When results are relevant, each video tends to cover parts of the topic instead of giving a fully comprehensive view of the whole topic. Users therefore lack guidance of which videos to watch and in what order: *"There were a lot of options to choose"*. Trying to identify the appropriate videos that fully cover the topic becomes a highly demanding task: *"Took me time to research"*. In fact, 26.4% of visits to watch videos for self-development were reported to also include visits to other search engines. Visits to YouTube for self-development were reported to be somewhat less urgent for lower-engaged users (36.0%) than for highly-engaged users (47.0%, $\chi^2 = 9.9, df = 1, p < .01$).

Get Instructions to Do Something

In 35.8% of all information visits, participants reported looking for steps on how to do something, most commonly by watching how-to videos. Intent for these motivations (69.5%) was more clear than for self-development motivations (58.6%, $\chi^2 = 12.78, df = 1, p < .01$), and best portrayed the instrumental use of online video as described by previous work [3]. This is further exemplified by one participant's report: *"I was looking for an instructional on Excel 2010 on how to do v-lookups"*. Similar to other informational motivations, participants were concerned about accuracy of the information: *"Lots of videos. It's hard to know who is qualified to give good information"*. In fact, reports of use of search engines along with watching YouTube were quite high (33.5%), which could be related to users double checking information or expanding their research. When finding the right video, often participants

considered it challenging to move between steps: *"There were long videos I had to skim through. I was wishing that a lot of these videos had chapters I could easily see the content and jump to"*. Ads experience needs to be carefully crafted as they get in the way of solving a problem: *"The commercials were distracting"*. Note that urgency of addressing this motivation is amongst the highest across all motivations (62.6%), although it was significantly less urgent for lower-engaged users (43.3%, $\chi^2 = 29.75, df = 1, p < .001$). High urgency could have been one of the reasons why participants were more sensitive to interruptions from ads.

Make a Decision

Of all informational visits, 15.3% were to get support for making a decision, most often related to a purchase (similar to product review motivations identified in previous work [7]) *"I was looking for makeup reviews because I was thinking about purchasing it"*. Users also reported using search engines alongside YouTube relatively often (35.8%), which could be related to getting accurate and further information for the decision. It's a process that takes time: *"I had to look at a few videos before I found the one that helped me"*. One of the main challenges was to distinguish reviews made by manufacturers versus customers, e.g. *"I wanted reviews from non biased people but could only find representatives"*. And finally, participants were concerned when videos were not fresh enough to support the decision making: *"Not enough recent videos, mostly dating back a year to years"*.

Get the Latest information

Watching videos to get the latest updates on a certain topic occurred in 7.1% of informational needs. Common content categories included news, entertainment, and video blogs, such as in this report: *"I wanted the morning news so I could hear and see what's going on in the world"*. Once again, other search engines were the primary tools used alongside YouTube (36.6%), indicating that informational needs tend to be mainly addressed using search. A key challenge was to find the most fresh content about a topic: *"It did not have the most recent stories available"*. Finally, participants reported the need to get updates within a specific context: *"[Could not find] immediate weather updates specifically for my area"*.

Motivation of Connection

The motivation to connect with others accounted for 9.8% of visits to YouTube using mobile devices. They were mostly pursued with at least one other person (62.2%). The relatively low frequency of this motivation was consistent with previous research that found needs for companionship to be uncommon, especially for online video [3]. Although relatively infrequent, these motivations were overall the most urgent needs to be fulfilled (63.2% were classified as urgent, compared to 52.5% for entertainment and 47.6% for information, $\chi^2 = 53.92, df = 2, p < .01$). Motivations for connection were classified based on the direction of the pursued connection: from the respondent to another person was labeled as a "proactive" connection, from another person to the respondent was labeled as a "reactive" connection, and connection that went both directions was labeled a "mutual" connection (See

Table 5). In the following subsections, we give details on each of these connection motivations addressed with mobile devices.

Motivation	% Occurrence	User State of Mind	
		% Clear intent	% Urgent
Proactively connect	49.1%	68.1%	67.0%
React to connection	42.2%	64.1%	57.5%
Mutually connect	7.7%	63.2%	70.2%

Table 5. Description of Connection motivations. Occurrences are percentages of all Connection motivations addressed with mobile devices.

Proactively connect

When trying to connect with others, participants most often did it proactively (49.1%). There were two ways that proactive connection motivations presented themselves. The first one was when the respondent created a need for someone else in order to connect socially, e.g. *"[Looking for a] specific movie trailer I wanted to show my girlfriend"*. The second was when participants learned someone's need and proactively tried to address it, e.g. *"My parents rescued some kittens and I was trying to find some help information to give them"*. Proactive connection motivations often led to a co-watching experience. The majority of proactive connection cases involved two persons (47.5%) in the video watching experience, or a group of people (18.7% cases with more than two people). The key reported challenge was video sharing with peers. Sometimes it's important to curate and/or edit content before sharing it, e.g. *"I found the video but there's no way to cut just the part that I wanted to share"*. In other cases, it's a matter of task efficiency: *"[The challenge was] getting the link, copying it, and sending before my friends moved on in the conversation"*. When using other tools alongside YouTube, search engines and social networks were the most common tools used (19.8% and 13.7% respectively). Social networks may have provided triggers to send videos to peers, while YouTube was the platform used to expand the social experience. Although this motivation was typically quite urgent, it was less urgent for lower-engaged users (61.4%) than for highly-engaged users (73.1%, $\chi^2 = 5.17, df = 1, p = .02$).

React to connection

The second most usual form of connection motivation was to watch or look for videos by reacting to someone's personal suggestion or request. Reactive motivations accounted for cases when peers had more important needs and depended on the participant's reaction, e.g. *"My daughter wanted me to make a flower on her bracelet and I needed a tutorial"*. Participants most often addressed reactive connection motivations in a co-watching experience with another person (40.0%) or a group of people (13.4%), but were alone relatively often (35.6%) as well, like in this example: *"My friend sent me a video, so I watched it."* Entertainment, music, and video blogs were the most common categories of video watched. When YouTube was used with other tools, social networks were the primary one (18.4%). In these cases, social networks often triggered peers' suggestions and requests to participants, who then reacted to them. Since these requests often came from

kids to their parents, looking for age-appropriate content with high-educational quality was an important need: *"It's hard to find quality kids content"*. Finally, the need to use multiple tools imposed even more challenges: *"When I clicked on the video [on the social network], it didn't open in the app so I manually searched for it"*.

Mutually connect

In 7.7% of connection cases, participants reported watching videos to bond with someone else or with a group of people who shared the same need, e.g. *"We wanted home improvement shows because my girlfriend and I like watching them"*. Participants were rarely alone (10.5%); mutual connection motivations were mostly addressed in a co-watching experience with a group (43.9%) or a pair (42.1%). In agreement with related work [16], we observed that negotiation on a video to watch can take significant time and be detrimental to the experience, e.g. *"Took too long and we lost interest"*. The inefficiency seems to be caused by the lack of personalized recommendations to groups of people, e.g. *"[challenging to] find our rhythm of songs we would enjoy"*. For mutual connection, entertainment and animation were the most commonly watched genres. Other tools such as search engines (10.5%) and social networks (5.3%) were almost never used along with YouTube.

Comparison to Non-mobile Devices

We observed some interesting differences in frequencies of device usage across motivations for watching videos. For example, traditional non-mobile devices such as TVs were more often used than mobile devices for motivations of social connection. The key reason seems to be that co-watching was more prevalent for TV use than for mobile device use (54.7% and 18.0% respectively, $\chi^2 = 104.96, df = 1, p < .001$). Some shared environments encouraged a regular co-watching habit, e.g. *"[it's] part of our bedtime routine."* Serendipity also played an important role for households, who would naturally join a video watch experience already happening on a TV, as it can be noticed in this user's report: *"I just joined my dad who was watching TV. Did not care about what we watched."* Indeed, scholars have presented several challenges for co-watching videos [16], and the private nature of mobile devices seem to play an important role in these challenges. HCI practitioners should feel encouraged to design tools that leverage the pros of personal phones better understanding each one's interests, and using that to facilitate pleasant co-watching experiences for its participants.

On the other hand, addressing motivations of information had minor differences in terms of device usage. Self-development was slightly more common motivation when using a computer than when using a mobile device (16.4% vs. 12.8% respectively, $\chi^2 = 11.08, df = 1, p < .01$). The work context seems to be relevant for addressing these motivations using computers, as in this example *"I was actually instructed to view videos for training purposes [...] for work"*. Schools also played an important role for information needs accomplished in computers, e.g. *"[Needed] help on a class I'm taking [to] learn about accounting."* Still, the practical difference in usage

frequency between computers and mobile devices might not be considered significant, which shows how relevant mobile technologies are becoming as learning tools.

Finally, we observed a difference in the high frequency of these motivations when televisions were used (71.1%) compared to mobile devices (59.5%, $\chi^2 = 6.47, df = 1, p = .01$). Narrowing down the analyses to specific entertainment motivations revealed some interesting patterns. For example, in a previous section we discussed how context can play a role in the decision of which device to use, such as the higher usage of non-mobile devices versus mobile devices to focus on a certain task *at work* (48.5% versus 12.9% respectively). It's important to notice that usage of TV in our dataset is related only to watching YouTube videos, and thus it excludes the very usual scenario of killing time by watching cable TV or TV on-demand, as extensively covered in prior work [13, 3].

IMPLICATIONS FOR DESIGN

The framework presented in the previous section contributes to a better understanding of people's motivations to watch videos (in-depth framework), their state of mind (clearness of intent and urgency), their context (co-watching or alone), their behavior (frequency of addressing motivations and other tools used alongside YouTube), and key challenges to fulfill their needs. In this section, we present a summary of 9 key recommendations for design in order to improve user's experience with their video watching journey.

1. Provide guidance, planning, and better curation: When users wanted to learn more about broader topics for self-development, they lacked a proper structure to guide them in the learning process. Enabling video creators to curate content towards more defined structures could address this challenge, e.g. supporting creation of modules with sequences of parts of videos to cover a macro topic. In addition, providing goals for curated educational content and expected time to be invested could support user's planning, a key challenge they faced when addressing their informational motivations.

2. Highlight legitimate sources and enable credential verification: In a world where anyone can make and share videos online, highlighting trusted sources and content providers could address user's need for choosing the right content when they were motivated to watch videos for getting information. Also, it's important to create a credential verification process that allows new content providers to become established. More research is needed to identify an appropriate verification method that is accurate and can be scaled.

3. Provide and highlight different kinds of "fresh" content depending on the user motivation: We observed two interpretations of fresh content: videos that were recently made and videos that were made about recent events. The former was important across all needs, and particularly for entertainment (e.g. watching latest content from a video creator the user has a lot of interest in), while the latter was key to informational needs (e.g. looking for news content). Video sharing websites need to better identify and highlight videos made about recent events. Highlighting could be done by calling out fresh content in news search results, e.g. in a separate

results group and/or prioritizing fresh content in the results list depending on how informational the search query is.

4. Distinguish reviews by manufacturers and customers, especially useful for information motivations: When deciding whether to buy a product or service, users usually wanted to learn more about it from two sources: the manufacturer was trusted to explain details about the product/service and customers were trusted to provide an unbiased evaluation for it. Video search engines need to cluster these two groups of videos separately when listing search results for any given product or service to better support user's decision making.

5. Ease discovery with diverse content in search results, particularly for motivations related to emotional states: Particularly when looking for entertaining videos to be on a certain emotional state, users often needed to discover more diverse content, e.g. to avoid repeating the same song over and over again when trying to focus on a task, relax, have a laugh, etc. Video search engines should ease access to diverse content for any given search query, e.g. with clusters that trigger to interesting content to be discovered that week.

6. Minimize interaction for emotion-based motivations: For example, when watching or listening to videos to augment another task, users are looking for a lean-back experience in which they do not have to take any decisions for a while. It's important that video search engines identify context and search queries related to the *Focus* motivation and provide results that are not only compiled for longer sessions without need for further user interaction (as currently accomplished with video playlists), but also support a diverse set of results, e.g. by varying playlist content in future sessions, providing triggers to quickly change premises of videos what should be played next, etc.

7. Steer intent clarification through content consumption, especially when mobile devices are used: Broad queries like "funny" and "star trek" indicated that the user did not know exactly what they wanted (e.g. unclear whether the search "star trek" was for the movie, or the TV series, or the cartoon, or the toys, or interviews with the cast, etc.) This kind of search was frequent in mobile devices because of text input limitations. In these cases, we recommend video search to act as a guiding tool that helps users disambiguate their intent through experiencing content, e.g. deciding which branch of the broad search query to go deeper based on clustered videos and their metadata, providing short video previews in search results to steer decisions to explore certain subsets of results, etc. As a secondary approach, search disambiguation could be addressed beyond search auto-completion using meaningful pivots to subsets of themes.

8. Identify co-searching and co-watching scenarios, and better tailor recommendations for the group: Video recommendations are usually based on interactions from a single user. In co-watching scenarios, that becomes challenging since recommendations will likely please the user conducting the search on their own phone, but not necessary adjusting to preferences of the group. This leads to time waste: *"Trying to find something my husband and I both liked."* If

co-watching scenarios are accurately identified, video search engines should combine and optimize video recommendations for groups of users in an attempt to address the co-watching challenges observed in our study and in prior literature [16]. We observed that these scenarios often follow a pattern, like in this example: *"My husband and I watch it together. This has been everyday."* We suggest exploring the pros and cons of different approaches, such as automatic context identification using sensors, detection of interaction patterns related to groups of users, support for self-reported group experiences (e.g. through account switcher), etc.

9. Better support sharing and social interactions: Motivations to connect with others through videos rely on good models for sharing content. This was challenged by participants in our study who had to come up with some workaround, e.g. *"Sometimes we'll be on our different phones doing our own search on the same topic. We'll switch between phones to show the other what we found. We do that hundreds of times per day."* We encourage more work on video sharing solutions to address these problems, e.g. effectively supporting co-watching, dialogs between peers, negotiation of which videos to watch, interruptions in the watching experience, among other dynamics of social interactions.

CONCLUSIONS

Our methodology enabled the creation of an in-depth framework on user's motivations to watch videos on YouTube when using mobile devices. In this framework, we describe details about the 14 identified motivations, particularly about the user's state of mind (clarity of intent, urgency of addressing the need), context (co-watching or not), behavior (frequency and tools used) and challenges to fulfil related needs. The method was described in detail to support reuse and replicability. We encourage reuse of this method by expanding on the probing surveys (see Table 1) in further studies in order to explore and report on other dimensions not discussed herein, as well as to explore design solutions across different video watching motivations.

ACKNOWLEDGEMENTS

We would like to thank Elizabeth Roodhouse, Kerwell Liao, Natalie Johnston, Mrinal Sharma, Regina DeNatale, Heather Arnold, Bob Evans, Laura Naylor, and all participants of our studies.

REFERENCES

1. Lisa Feldman Barrett and Daniel J Barrett. 2001. An introduction to computerized experience sampling in psychology. *Social Science Computer Review* 19, 2 (2001), 175–185.
2. Kathy K Baxter, Anna Avrekh, and Bob Evans. 2015. Using Experience Sampling Methodology to Collect Deep Data About Your Users. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 2489–2490.
3. Beverly A Bondad-Brown, Ronald E Rice, and Katy E Pearce. 2012. Influences on TV viewing and online user-shared video use: Demographics, generations,

contextual age, media use, motivations, and audience activity. *Journal of Broadcasting & Electronic Media* 56, 4 (2012), 471–493.

4. David Geerts, Ishan Vaishnavi, Rufael Mekuria, Oskar Van Deventer, and Pablo Cesar. 2011. Are we in sync?: synchronization requirements for watching online video together.. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 311–314.
5. Giovanni Iachello, Khai N Truong, Gregory D Abowd, Gillian R Hayes, and Molly Stevens. 2006. Prototyping and sampling experience to evaluate ubiquitous computing privacy in the real world. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*. ACM, 1009–1018.
6. Stephen S Intille, John Rondoni, Charles Kukla, Isabel Ancona, and Ling Bao. 2003. A context-aware experience sampling tool. In *CHI'03 extended abstracts on Human factors in computing systems*. ACM, 972–973.
7. Christoph Lager, Mathias Lux, and Oge Marques. 2012. What makes people watch online videos: An exploratory study. *Computer Entertainment* (2012).
8. Jennifer McNally and Beth Harrington. 2017. How Millennials and Teens Consume Mobile Video.. In *Proceedings of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video*. ACM, 31–39.
9. Jessica Gall Myrick. 2015. Emotion regulation, procrastination, and watching cat videos online: Who watches Internet cats, why, and to what effect? *Computers in Human Behavior* 52 (2015), 168–176.
10. Daisuke Ochi, Hideaki Kimata, Hajime Noto, and Akira Kojima. 2012. Scenario-driven interactive panorama video delivery: promptly watch and share enjoyable parts of an event. In *Proceedings of the 20th ACM international conference on Multimedia*. ACM, 1307–1308.
11. Kenton O'Hara, April Slayden Mitchell, and Alex Vorbau. 2007. Consuming video on mobile devices. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 857–866.
12. Dante M Pirouz, Allison R Johnson, Matthew Thomson, and Raymond Pirouz. 2015. Creating online videos that engage viewers. *MIT Sloan Management Review* 56, 4 (2015), 83.
13. Steven Reiss and James Wiltz. 2004. Why people watch reality TV. *Media psychology* 6, 4 (2004), 363–378.
14. J. M. Rigby, D. P. Brumby, S. J. Gould, and A. L. Cox. 2017. Media Multitasking at Home: A Video Observation Study of Concurrent TV and Mobile Device Usage. In *Proceedings of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video*. ACM, 3–10.
15. A. Shokrpour and M. J. Darnell. 2017. How People Multitask While Watching TV. In *Proceedings of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video*. ACM, 11–19.
16. Emily Sun, Rodrigo de Oliveira, and Joshua Lewandowski. 2017. Challenges on the Journey to Co-Watching YouTube. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, 783–793.
17. Yi Zhu, Ingrid Heynderickx, and Judith A Redi. 2015. Understanding the role of social context and user factors in video Quality of Experience. *Computers in Human Behavior* 49 (2015), 412–426.