

# Challenges on the Journey to Co-Watching YouTube

Emily Sun  
Cornell Tech  
New York, NY  
es765@cornell.edu

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

Joshua Lewandowski  
YouTube  
San Bruno, CA  
lewandoj@google.com

## ABSTRACT

In order to better understand social aspects of the short-form video watching experience, we investigated the journey to co-watching, from searching and discovering content, to choosing and experiencing videos with others. After identifying, through a large-scale survey, some of the most typical situations that bring people to YouTube, we deployed a one week-long diary study with 12 participants in which they performed a set of frequent video tasks at their leisure, half by themselves, and half with someone else. Following the diary study, we had participants reenact the diary study tasks remotely with the experimenter. We observed that users face multiple challenges on the journey to co-watching a video. They must share a device designed for an individual, use different methods for selecting videos than when by themselves, negotiate or turn-take in order to make a decision, and potentially watch a video that they do not enjoy. Along this journey, users must engage in impression management to consider how their choices might make them appear to others. We present design recommendations for remote and collocated co-watching to improve the social video watching experience.

## ACM Classification Keywords

H.5.2. User Interfaces

## Author Keywords

Social video; Co-watching; Co-discovery; YouTube

## INTRODUCTION

A recent Nielsen report showed that nearly two-thirds of people around the world watch some form of video-on-demand programming, 43% of whom say they watch at least once a day [5]. Social interactions and collocated watching are an integral part of enjoying videos. According to Macaranas et al., 75% of people report watching video content with one or more people daily or weekly [20]. People have conversations during and after the video watching experience that enhance the appreciation for the video, and researchers have even developed systems which explicitly support audio conversation with remote video watchers [7]. The importance of collocation

is reflected in recent CSCW research that has examined and designed systems for co-presence [1, 26, 35].

Scholars and HCI practitioners have created systems to support video watching even when not collocated. Recent developments in video systems have focused on synchronizing video playback for users watching remotely [9]. Sharing rich content during calls can enhance the enjoyment and sense of connectedness for remote communicators [8]. Systems such as Zync allow people to watch together at the same point in a video and maintain discussions when watching video at a distance from one another [31]. Other systems, like audience silhouettes, aim to increase social presence through representations of a remote viewer's body during co-watching [34]. Emotars, a combination emoticon and avatar that communicate feelings, have also been researched as a way of enhancing togetherness during remote viewing [14].

These systems focus only on enabling co-watching, but little is known about the user journey that leads up to co-watching, and how it should be best supported. In fact, the experience of watching a video together can start before users hit the play button. We aim to understand why people choose to co-watch a video, as well as how they decide which content to watch and share. After a video has been watched, it is unclear how the video watching experience continues in the presence of someone else. Previous research on TV watching in the home provides insight into how families dynamics influence co-watching, but further work is needed to understand co-watching of short-form content, which can be viewed in a greater variety of contexts and fluid situations. Situational factors like where users watch [25], how many screens are present [3], and who is present [24] could have a significant influence on how people interact with a video and each other prior to and after co-watching. We focused on collocated shared viewing since this is an underexplored area in research. Still, our results could inform the design of both collocated and remote video watching experiences. To summarize, our research aimed to understand:

- What are the steps that viewers take to co-watch?
- In what contexts does co-watching occur?
- How might we design systems to better support co-watching?

In this paper, we present key challenges in the journey to co-watching videos, from (1) searching and discovering content to (2) choosing and experiencing a video with others. We look at situations that cause social sharing to occur, as well as how

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. Copyright is held by the owner/author(s).  
CSCW'17, February 25 - March 01, 2017, Portland, OR, USA.  
ACM 978-1-4503-4335-0/17/03.  
<http://dx.doi.org/10.1145/2998181.2998228>

this sharing takes place. In order to study co-watching, we deployed a large-scale survey, diary study, and reenactment sessions. We found that if people are motivated to co-watch and select a device to watch on, they spend more time searching for a video when co-watching, engage in negotiating or turn-taking to decide what to watch, and may watch a video that no one is enjoying. Finally, we present design recommendations on how to improve the collocated watching experience and how remote systems could better support the social experience of watching videos.

## RELATED WORK

Previous literature has examined the social influence on video watching from a broad perspective. Mercer et al. detailed the contexts in which social watching occurs. Their results show that people watch videos socially in both public and private settings, and that this viewing can be engaged or unengaged. Engaged viewers are engrossed by the content, whereas unengaged viewing is primarily motivated by social reasons. They found that live events in particular were enhanced by the social element of group watching [24]. Browsing YouTube can also be “occasioning,” where a user may engage with the people around them, but return to the app at opportune moments. [2]

Relevant research can be found from the domain of television research. Co-watching on TV at home was found to depend on contextual factors. Different viewing situations included family quality time, relaxing after school, and lazy afternoons, each of which was associated with a different time, mood, content, and viewer [33]. Similar to unengaged viewing, Ley et al. reported that people in a household would co-watch TV but with “different interests” [16]. In this situation, people would use their smartphones separately while remaining in the presence of the person whose main attention was on the TV. In a large-scale study analyzing logged household video watching, researchers found that group watching and individual watching differ significantly by genre. Chaney et al. developed a model to determine a group’s viewing interest based on those of the individuals [4]. Their work revealed that co-watching is more than a simple aggregate of individual preferences, but the scope of their research did not delve into *why* these differences exist.

Related is also work by McGill et al. [23] that examined different ways a single-user interface for control of a TV supports multi-user use. They proposed different control schemes for how people might share control, divided into “one user schemes,” where people shared one remote interface, and “everyone schemes”, where control was divided amongst the co-watchers. Their survey results suggested that “one user schemes” perform better than “everyone in control,” though it is unclear how much of their participants’ responses might be due to familiarity with current systems of use. Other research has explored how family roles and communication patterns affect television selection [19, 18] and worked on developing models to aid in group-decision making for TV co-watching [7, 21]. While useful to build upon, TV research may not directly apply to short-form video content since co-watching of YouTube could occur in a greater variety of fluid contexts with different people.

Outside of the home setting, O’Hara et al. reported that while participants cited “passing time” as the reason for their viewing, they often were using video to manage relationships with others. Examples they gave include watching video to avoid interacting with other people in a car or disengaging while in public space [27]. Hampton & Gupta [12] categorize “place-maker” as those who use digital content to create connections with collocated people in public spaces. The social influence of watching in a public setting leads to changes in people’s behaviors. Culture may have an influence on social watching; a survey of Australian attitudes towards mobile video showed that 7.5% of users watch for the purpose of sharing. [32] Miyauchi et al. reported that Japanese mobile TV viewers stop watching in social settings because of social disapproval, privacy, and lack of appropriate content [25]. However, they also found that video can be used as a social lubricant; friends at a cafe view video together as a way to help continue a conversation.

Building off of Goffman’s work, Lin et al. identified that impression management is a factor in people’s mobile video usage [17]. They showed that watching mobile video may be used as a self-presentation strategy. In the same way that self-presentation is a concern for users of social media like Snapchat [36] and Facebook [30], Lin et al. found that making other people aware of their usage of mobile video and taking pride in the action was one of the factors that influenced people’s intent on watching video. Similarly, Matthews et al. showed that people use their mobile phones to broadcast video content with people they were with as a way of sharing [22].

The research landscape provides a broad description of the kinds of contexts in which co-watching occurs, the social influence on when and what to watch, and co-watching of TV inside the home. In this study, we aimed to complement previous research by conducting a micro-examination of the social watching situations of short-form content, in a variety of different contexts, to gain a deeper understanding of the steps that people take to co-watch and the main challenges faced in this process. While general co-watching situations and influence of others have been studied, specifics on how one can influence strategies to find and consume mobile video in a co-watching experience have not been fully explored. Following, we present the study conducted to gather insights into how people try to satisfy their video needs in the presence of others.

## METHODS

We focused our study of video co-watching on YouTube, the world’s largest<sup>1</sup> video sharing platform with over a billion users and hundreds of millions of hours watched every day. We employed three methods to study co-watching: large-scale survey, task-based diary study, and reenactment sessions.

With the large-scale survey, we reached out to over 3,000 people, aiming to reveal some of the most typical tasks around watching videos on YouTube. This was important to generate a list of common video watching tasks.

<sup>1</sup><https://www.youtube.com/yt/press/statistics.html>

Participant	Gender	Age Range	Occupation	Self-reported YouTube Usage
P1	F	31-40	Operation Manager	Multiple times a day
P2	F	24-30	Newscast Director	Multiple times a day
P3	F	24-30	Call Center Representative	A few times a week
P4	M	31-40	Business Analyst	Multiple times a day
P5	F	31-40	Wedding Magazine Editor	Multiple times a day
P6	M	31-40	Student	Multiple times a day
P7	F	41-50	Plan Sponsor Liaison	About once a day
P8	M	24-30	Content Technologist	Multiple times a day
P9	M	24-30	Press Operator	Multiple times a day
P11	M	18-23	Finance Center Intern	About Once a day
P11	F	31-40	Teacher	Multiple times a day
P12	M	41-50	Sales Representative	Multiple times a day

Table 1. Participant Demographics

Next, we conducted a diary study with 12 users living in the US, and collected detailed information about instances when they performed typical video watching tasks—identified by the large-scale survey—either by themselves or with someone else. The diary study allowed participants to report on their behavior in their natural environment, rather than in an artificial lab setting. However, to ensure that we would obtain enough information about co-watching in situations that people commonly use YouTube, we assigned several tasks for users to perform during the week of their diary study.

The last phase of the study was reenactment sessions that included semi-structured interviews with the diary study participants. During these video conference sessions with the researcher, participants mimicked the steps they took in the diary study tasks, using their diary entries to jog their memories. Having participants re-enact their process encouraged reminiscing and triggered details that had been left out of their diary entries [6]. The reenactment sessions of the diary tasks allowed us to learn (1) more details about their experience in tasks that they regularly engage in co-watching in, as well as (2) reflections on their attempt to co-watch in tasks they are not accustomed to. From the diary responses and reenactment sessions, we identified which tasks would fall on either of these cases, and took this information into account when analyzing the data to generate insights. For example, occurrences of the first case helped us to investigate how the journey to co-watching often happens, whereas occurrences of the second case were useful to reveal assumptions for why people choose not to co-watch and to learn from first-time co-watching experiences for any given task. Participants received monetary incentives to be part of the diary study and the reenactment sessions.

### Survey to Determine Tasks

In order to inform the design of our task-based diary study, we ran 18 Google Consumer Surveys<sup>2</sup> with 200 participants each, asking about their motivations for using YouTube. Multiple surveys were used to allow asking only a few of questions in a given survey to reduce drop outs, while collecting answers

to multiple questions with multiple surveys. Each survey consisted of three questions: an attentiveness check, a motivation question and a context question. The attentiveness check was included to ensure the quality of responses, as recommended by Goodman et al. [11]. The motivation questions were open-ended, asking users one of the following: (1) what brought them to the *last* video they watched, (2) why they *usually* visit YouTube, or (3) what led them to an *important* video that they watched. The context questions asked participants to describe the situation they were in when they watched the video, the time of day, location, multi-tasking, devices, and who else was around them. One question from each block was presented in each survey, for a total of three questions per survey.

The results of the survey were analyzed broadly to hone in on the tasks to assign for the diary study. Music was the most listed reason for watching YouTube. Participants described their motivators including moods like being bored, depressed, and curious, and the desire to learn something like fixing an appliance or needing a new hairstyle. Half of the respondents described watching YouTube mostly at home (50%) and some reported multi-tasking while watching (10%). Most people watched alone (65% across all surveys), but 24% of the survey respondents who were shown the “important” motivation video question, as described above, responded that they watched it together with someone else. Laptop/Desktop (55%) and mobile devices (52%) were the most popular devices, followed by tablets (36%). In the open-ended responses, people included reasons for watching YouTube like:

*“I was sitting down with my mom on her tablet and were crowded around looking up the video of my niece experiencing snow.”*

*“Son with autism having first MRI and we wanted to let him hear the sounds.”*

The survey provided us with a broad understanding of when and why people visit YouTube. The high reports of mobile usage and video-watching outside of the home indicated that watching behavior of short-form content like YouTube might differ from previous research on TV-watching. Using this data, we developed and iterated on the tasks for the diary study.

<sup>2</sup><http://www.google.com/insights/consumersurveys>

Entertainment Tasks	Information Seeking Tasks
Find a video to entertain yourself for a couple minutes.	Find a video with step-by-step instructions on how to do something.
Find a short, funny video.	Find a video related to one of your hobbies.
Find a mellow song to play.	Find review videos to help you make a product decision.
Find something to play if you wanted to workout.	Identify a movie you would want to see in theaters.

Table 2. Diary Study Tasks

### Participants

We recruited 12 YouTube users ages 18-50 for the diary study from Google’s external participant database (see Table 1). The database is comprised of individuals who have opted<sup>3</sup> into receiving invitations for various Google/YouTube usability research studies. Google and YouTube use multiple strategies to populate and maintain this pool of participants, including in-product advertisements, newspaper ad spots, business cards, among others. These various approaches ensure a diverse and extensive pool of participants.

Half of the participants were male. Care was taken to select participants from across different parts of the United States. The participants were selected based on their responses to a screener. They were required to have access to hi-speed Internet in order to participate in the interviews via remote video call, and reported using YouTube at least a few times a week.

Participants reported regularly completing all the tasks from Table 2 between once a month and at least once a week. They also reported a range of completing between 2 and 8 of these tasks often with someone else (i.e. co-watching). For the diary study, each of the co-watching tasks had *at least* two participants who reported to often co-watch for that task. Their family situations differed from living alone to having children to living with their parents. Occupations were varied, including a newscast director, call center representative, teacher, and finance intern, among others. One of the participants did not complete all of the diary study tasks and did not proceed to the next phase of the study.

### Task-Based Diary Study

Based on the results of the large-scale survey, we crafted the eight tasks that participants were assigned to complete in one week during the diary study (see Table 2). Four of these tasks were in the entertainment category, and four of them were information-seeking tasks. They were familiar and used to performing the eight tasks as reported in their screener survey. Participants were instructed to complete the tasks as naturally as possible. Half of the tasks were randomly assigned to be completed alone, and half were to be completed with someone else of the participant’s choosing. Participants answered questions about the context they were in when completing the task (e.g. “What time of day did you complete this activity?”

<sup>3</sup><http://www.google.com/usability>

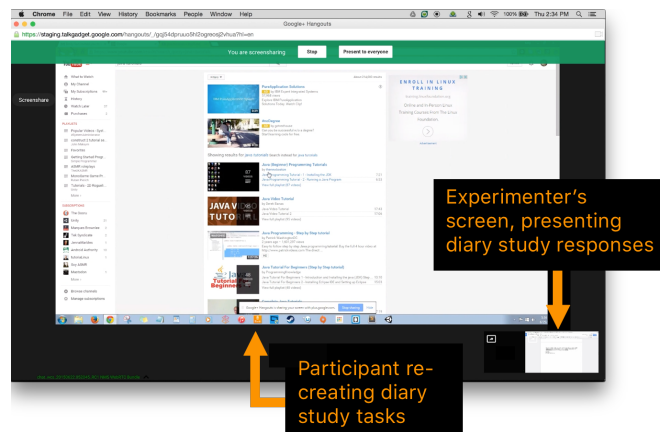


Figure 1. Screenshot of a reenactment session with the participant sharing his screen and explaining how he completed the task.

“Roughly how long did it take you to find a video?”) and the path that they took to arrive at their selected video for each of the tasks (e.g. “How did you navigate to the video?”). In order to gauge how reflective the tasks were of frequent usage of YouTube, we included a question that asked, “How similar was this activity to ways you actually use YouTube?”

### Reenactment Sessions

In the following week, the remaining 11 participants who had completed all the diary study tasks were interviewed by the researchers using Google Hangouts<sup>4</sup>, a video chat service. (see Figure 1) Participants were shown their responses from the week before to help remind them of how they had completed the tasks. They used the remote video call capability of sharing their monitor screen with the researchers, and re-created the video searching process from the diary study, talking aloud while performing the task. Completing this process for each of the tasks and answering follow-up questions resulted in 1.5-hour interviews.

Once the interviews were complete, we used a grounded theory method of analysis by coding recurring themes among responses in an iterative approach until we arrived at the ones described.

### Reflections on Methods

By asking any given participant to perform certain tasks with someone else, we arrived at two cases: either (1) the participant had experience co-watching for a particular task, or (2) the participant never tried completing those tasks with someone else before the diary study. The first case was key for us to understand the journey to co-watching videos. Based on their responses to the screener, we knew that case 1 would be found for all of the tasks for two to nine of the participants. The second case was important to reflect on obstacles to co-watching. By asking a participant to co-watch for a certain task that s/he never experienced through co-watching before, the participant was able to reflect on the reasons why s/he never did. Not in spite of, but because of the direct contrast in performing

<sup>4</sup><https://hangouts.google.com>

the task alone or with someone else, participants were able to speak to the differences between the two experiences of watching alone and co-watching.

For example, even though P4 often watches video with his wife, finding a video of a hobby with his wife was unnatural for him. In his diary study response, he wrote, *“I normally don’t complete this activity with another person (my wife included), as she mostly has separate hobbies from mine. So while this is something I use YouTube for often, it is not something that is often done with another person. We both felt a bit strange doing so.”* During the subsequent interview, he explained that even though his wife and him both share the hobby of travel, he predominantly does the research for their trips. Further probing in the interview revealed that *“In this instance if it was related to travel, I would use YouTube to search for places we would travel and things that would look interesting for us to do together and I would tell her after the fact...I would be doing it separate[ly] and then tell her about it afterward.”* From this exchange, we learn that overlapping interests is not the sole indicator of co-watching; social dynamics also play a key part in how people choose what to co-watch. The interviews may not have yielded as relevant conversations had the first phase of the study not required both social and individual viewing.

Similarly, having their diary entries presented to them and recreating the steps incited richer descriptions of the situations participants were in when they completed the tasks. Once participants read their entries and typed in a search term, they often responded with comments like *“Yes, this one I definitely remember.”* (P2) or *“Oh yeah, here’s what I did. I went to the Music tab...”* (P11). Interview questions were interspersed throughout the reenactment sessions to gain a deeper understanding of the motivation and context of participants’ actions. In large part, the diary tasks acted as a starting point for people to reflect on their everyday experiences.

## RESULTS

The diary study entries revealed that tasks were finished relatively quickly and with ease. Participants were able to locate a video within 0 to 2 minutes for 71.9% of tasks and found their video within 2 to 4 minutes for 20.2% of tasks. On a Likert-scale from 1(Easy) to 7 (Difficult), the average task difficulty was rated a 1.3. The tasks were broad enough that people were able to complete them in different ways. For the entertainment category of tasks, participants reported relaxing (after work, before bed), filling time (on commute, during lunch break), and getting energized (dance parties, working out) as the main motivators for their video choices. For the information-seeking tasks, participants reported watching videos to help with planning (camping, things to do in San Diego, movie trailers), learning specifics about a hobby (surfing, photography), and seeing other people’s experiences (product decision). The majority of responses reported completing the tasks at home (78.7%) and in the afternoon or evening (93.3%), though participants also reported completing them in the car, at a park, and at a coffee shop.

We present the journey to co-watching, whose steps include motivation for co-watching, sharing a device, searching for a

video, decision making, and watching the video (see Figure 2). Certain steps can be skipped depending on the co-watching situation, and while motivation for co-watching was often a starting point, participants described instances when they started co-watching when one of the viewers was already watching a video. Unless designated otherwise, the results presented below are from analysis of the interviews during the reenactment sessions, which built upon and clarified participants’ diary responses.

## Motivation for Co-Watching

We observed that the journey of co-watching can start from the user’s motivation to watch videos together with others. Next we present results about: (1) how people can have different levels of involvement when co-watching; (2) the influence of others to motivate co-watching even when not physically present; and (3) how this motivation can happen spontaneously or be part of a routine.

Participants interpreted “watching together” to indicate different levels of involvement from the other person. The range extended from passively influencing video choice to actively engaging in the video searching process. People can play a more passive role in the video selection but even influence the selection process by their mere presence. During the interview, P2 reported often listening to music from a video player while her family members are in the same room. The others do not contribute to the video selection, but she still considers them in her decision-making, opting to pick songs that are only in English. P2 described, *“Even if it’s just me choosing, it’s me trying to choose something that they won’t reject rather than just picking something for myself. There’s always more freedom because you know yourself better than you know anybody else.”* In this case, her family members were not engaging in watching the video at all, but their presence affected P2’s choice (see “Decision Making”). Similarly, P11 is a school teacher and likes to find videos to play for her class when she is in a park, however being in a public setting makes her more deliberate about the content playing. This echoes previous literature on video content selection in public places [25].

In other cases, the video was the focal point of conversation. In her interview, P7 described how she uses YouTube as a reference when it fits into a discussion. *“Normally I use YouTube to search for specific clips from movies or TV shows or to look for songs. It’s pretty much always something that I want to show to someone else. Like we were having a conversation and I made a reference to a dialogue line or a scene in a show or something like that, and they don’t get the reference so I want to look it up and show it to them.”* Even when the other person isn’t physically present, they could make a difference in an individual’s choice. P3 explained, *“I had times where I’m waiting for a friend to come over and I don’t want to get started doing anything, then I’ll just try and find something to be like ‘hey look what I just watched.’”* Here, the video is being watched to fill time [2], but with the explicit intention of sharing it with another person. A similar situation was described by P9, where he re-watches videos in order to share them with others. *“I usually watch the video before and*

*then I get so excited about it I've gotta show it to them [his co-workers] as well" (P9).*

For some users, co-watching is a regular part of their schedule. P4 and his wife have a daily routine of sitting side-by-side on bed surfing the internet together. In his diary entry, he reported finding a clip he had seen before from Saturday Night Live and played it on the television to share it with his wife. *"Every week or so we're sharing something to the Chromecast. That's pretty much how we watch most of our content."* The clips they share are often sourced from different social media channels since he predominantly uses Google+ and his wife is active on Facebook. P5 uses YouTube to listen to music every day as a way to unwind with her sister. *"We do have a 10-second dance party everyday at my house...Normally when everybody gets home so it's usually before dinner around 7, 7:30 or it's later 9:30 or 10. That's just our time together."* Her sister sends her videos throughout the day while her family members are at work, but they don't always get a chance to watch them. *"My day's so slammed, and my mom works during the day so when we get home it's, 'Did you see this? Did you see this? No, let me show you.' I want to say about 75% of the time we're watching it together. Even though she's already watched it, she's going to show us and watch it again."* (P5)

### Sharing a Device

The second step in the journey to co-watching is selecting a device to watch on. Participants reported (1) co-watching on many different devices, depending on the situation, and (2) sometimes even switch between devices in the same viewing session.

From the diary responses, participants reported using a variety of devices to watch video, including mobile phones, computers, tablets, and TVs. Interviews revealed that sharing one screen like a mobile phone or tablet was a common use case among co-watchers, though the small screen was not always ideal for co-watching. P3 explained, *"Typically I would say we're in the same room and I would just 'Turn your head and watch this' but depending on what it is, I may send him a phone link to it. Text him the link to it so he can watch it on his phone."* However, in some cases, people would switch from one device to another. When his fiancée was watching a clip from a comedian, P9 described, *"It [the video] was already pulled up on her iPhone, and there was a group of other people who wanted to watch it, too, so we pulled it up on her Apple TV."* When listening to a long music video, P2 switched between different devices depending on where she was in the house. *"I started it in the bedroom with my dad and then I moved into the kitchen and listened to it with my mom while we were making dinner. I threw it on the Chromecast so that it would be on the big screen while I was folding clothes."* While in some cases the searching and watching experience is made more complex by requiring users to share the same device, in other cases multiple devices are needed to be used in shared turns, or content to be projected in a common shared screen. P3 recounted a story when she and her bridesmaids watched videos together the night before her wedding for 2.5 hours. *"My bridesmaids and I were in the hotel room, and they hooked a laptop up to the TV and we kinda searched for funny*

*videos and just watched random things, went to the suggested videos from there."*

### Searching for a Video

If a recommended video is selected, then co-watchers may skip the step in the journey of searching for a video. However, since recommended videos are based on private viewing history, co-watchers often resort to searching for videos. Next, we report findings on how searching takes longer when co-watching than when people watch by themselves because (1) co-watchers' interests do not always align, and (2) co-watchers have to reach a verbal agreement on which search terms to use.

Finding a video while co-watching required a search tactic that would allow people to find videos that would appeal to both parties. A common difficulty was picking a video that would match both people's topics of interest, since often times people had different preferences from their significant others or family. P2 explained, *"I'm usually pretty private when it comes to my video watching just cause our tastes are usually pretty different so it's hard to find people to watch with."* Sometimes one person's tastes will overlap with the other person, but this may not always be the case. For P3, they both enjoy certain videos but not the ones that her husband typically watches. *"Anytime he watches videos it's more like computer part reviews, stuff I'm not interested in, but I'll be like 'hey look at this funny thing I saw'" (P3).* As a result, participants would look for different videos that they think would appeal to the other person. When P8 watches exercise videos with his brother, he picks out different types of videos than he does by himself because they have alternate workout styles. *"A lot of times I work out by myself now that I have my own apartment. When my brother came over, it's like, 'Alright now I've gotta find something that he's going to like, too, or he's not going to want to do anything'" (P8).* P2 explained how she considers her co-watchers to filter out certain videos, *"I wouldn't watch any of the real funny or outlandish or cutesy stuff with my dad just cause he'd be like, 'What the hell are you watching? Turn it off.' My mother I'm a little more open with but I would probably steer more towards either 'Last Week Tonight' or cooking stuff with her. I wouldn't watch much of the tech stuff with her cause she would just be bored...it's channeling what I want to watch through the filter of what I think they'll tolerate."*

During their reflection on the difference between co-watching and solo-watching, all but one of the participants reported spending more time searching for videos when co-watching than when watching alone. Even though participants often reported, *"Usually what I'll do is look through the recommended things."* (P4) when by themselves, they used this tactic less when co-watching. P7 explained that when he searches for content to watch with his wife, *"The recommended videos tend to be ones more that I would want to watch, not that she would want to watch."* He continued on to say, *"If I'm searching for something for both of us to watch, a lot of things that I would be at least willing to watch the first 30 seconds of just to see if it was slightly interesting she wouldn't be all that interested in. So I tend to search more and try different things."* P4 described the different strategies he uses to watch videos

with his wife. *“We’ll see what there is Recommended. If we’re in the mood for a particular type of thing, then we do searches. Usually it’s something that we run across on Facebook that’s a YouTube link and she wants to show me or I want to show her so we do a search to see if we can find the one we were looking for.” (P4)*

The process of searching for a video on a single device can require verbal coordination between co-watchers. P11 described how she and her sister watched videos together on her phone. Even though she had control of the device, her sister told her which terms to put in the search bar. *“For funny videos, I would put in ‘funny video’ and she would put in ‘make me laugh’.”* During the interview, the participant revealed that she would not normally use these keywords to search, but were entered as a suggestion from her sister. While watching with her sister on a TV screen, P5 had a similar interaction but instead agreed on a search term to use. She described, *“I said ‘if we were working out, what would you want to do? Would you want something more upbeat and hip hop or something more low key?’ And she said, ‘hip hop.’ And I said, ‘well should I look up ‘hip hop’ or ‘upbeat’ and see what we get?’ And we both agreed on ‘upbeat.’”* In one instance, using different search terms lead to a participant being unable to find a video during a reenactment session. *“That’s probably why I can’t find the one that we watched because she probably told me what to say. Like I put in ‘surfing video’ and the one we watched didn’t come up so she probably said something else and I typed that in.” (P11)* Once search terms are input, users share the interface presenting the results. P12 described how it takes longer to look through the results when with someone else. *“If it’s me and my wife she wants me to slow down. She wants to see the description of each video.”* Since users focus on different meta-data when selecting a video, scrolling through results on one device can become more difficult when co-watching.

### Decision Making

A key step in the journey to co-watching is deciding on what video to watch. When more than one person is selecting the video, making a decision is non-trivial. We found that people generally employed two strategies to aid in the decision making process: (1) negotiating and (2) turn-taking. We also report our findings that people often co-watch pre-viewed content to minimize the burden of decision making.

Negotiating primarily has one driver of the device, who actively engages the other person in the choice. By asking the other person to verbalize his/her opinion, the driver decides if the video will appeal to all the viewers. This differs from the “everyone” scheme defined by McGill et al. [23] since one person is technically in control of the device, but that user is not the only one making the decision. This joint-decision making process can also lead to longer decision times. P3 explained, *“You have to wait for their feedback as far as, hey do you want to watch this? Does this appeal to you? Whereas when I’m by myself, I know exactly: I like it, I don’t [like it], and I’m onto the next one.”* The negotiating process can also be unequal, which may result in the driver dominating by selecting the video that he/she wants to watch. P2 explained

how even as the driver, she takes into consideration the co-watchers’ preferences. *“I’m controlling it and I’m sharing it with them but at the same time I’m very much restraining my choices because I know their taste.” (P2)* In some cases, this can be positive as happened with P5. *“I was holding the Chromebook at that point. She was sitting beside me looking at the computer screen, too. The video that I would pick I actually picked, but then it proved to be a positive thing for her, too, because that was song that she was humming that she didn’t know the name of.” (P5)* On the other hand, P5 explained that she and her sister have different tastes in videos and that they usually watch the video of P5’s choosing because she is the older sister. *“Cause she’s just as picky as I am, and we normally fight over what we’re going to watch. We have two different tastes, we’re totally different people...she’ll want to look at something from somebody else, and I’ll say no this looks better or they don’t look like they know what they’re doing...we argue and I win and we watch what I want to watch.” (P5)* This relationship dynamic reflects previous findings on family-role dominance in shared TV-watching [19].

Turn-taking involves each person acting as the driver at different times, a form of the “one user in control at a time” scheme denoted by McGill et al. [23]. P11 described turn-taking with her sister, *“She’ll choose something that she’s seen already and shows me, and then I’ll choose something I’ve seen already and show her.”* Users engage in a back-and-forth, handing off control after each video. This process is more equitable than negotiating, but also results in a user having to wait while the other finds and picks the video. P3 described a situation where turn-taking took place amongst a group. *“One person was controlling the computer, and it was four of us total. We kinda went, ‘Okay, you pick something.’ Then they’d pick something and then we’d all be like, ‘Look at that suggestion-that looks fun.’ or ‘Try searching this!’ It was taking turns but also we’re all agreeing on what to watch” (P3)*

As a result of the difficulties with decision making, users tend to select videos that they have seen before in order to minimize wait time. *“It’s usually something he’s already listened to and knows that I like or it’s something I found on YouTube the day before and it’s like, ‘Hey man, you have to check this out. It’s good stuff.’” (P8)* Participants reported waiting to watch certain videos until they were in the presence of the co-watcher. P9 described how a common situation when co-watching with his friends is, *“I found this trailer and I won’t watch it until I come over and show you as well so we can both see it together.”* By holding off on watching alone, participants are able to experience the enjoyment of seeing each other’s first time reactions. As part of their co-watching routine, P4 and his wife have a video that they only watch together, *“We have a show where I try not to watch it at all without her there.”*

### Watching

The last step of co-watching is experiencing the video together. Next, we present results on (1) how people may be unsure of whether a co-watcher is enjoying a video or not without verbal



**Figure 2. Steps on the journey to co-watching. Although it often starts with motivation, certain situations can have a different starting point or even skip steps.**

confirmation, which potentially results in (2) watching a video that neither party is actually enjoying.

Participants reported switching videos less frequently when watching with someone else rather than alone. P3 explained the interaction that occurs once a video starts playing that she's unsure if the co-watcher is enjoying. *“For the most part it's just like, ‘Hey you wanna watch this?’ or, ‘Hey let's try’ and then you get 10 seconds in and it's like, ‘Is it still good? Or just weird? Should we go back and try again?’”* (P3) A closer examination of a user's interaction hints at a reason why this behavior occurs. P8 reported playing YouTube music videos in his car with another person. He often tried to find music that both would enjoy, but would also listen to music that he would not listen to by himself to accommodate for the other person's taste. He said, *“If he [the person with the participant] likes it [the song] but I don't like it I'll still just listen to it.”* A potential psychological reason for this occurrence is pluralistic ignorance. Typically used to explain the pervasiveness of certain norms, pluralistic ignorance is “the case in which virtually every member of a group or society privately rejects a belief, opinion, or practice, yet believes that virtually every other member privately accepts it” [29]. In this situation, the user thinks that the other person is enjoying the video so he continues to watch it, when in reality neither person likes the video, but nobody wants to speak out against the other's perceived opinion. As a result, a pair or group of people could end up watching a video that nobody is enjoying. Since people are aware of their own opinions but

not necessarily those of the co-watcher, they are more hesitant to switch videos with others.

## DISCUSSION

The diary study and interviews revealed that users follow a complex journey to arrive at co-watching a video (see Figure 2). They often share a screen, search for content, find and decide on a video, and decide whether to keep watching videos or not. Each step requires balancing what the user wants and what could also satisfy the person engaged in co-watching. Even if the co-watcher is not actively engaged in watching the video, the person's preferences are influencing the watcher's choice. Co-watching is a routine for several families, where they sometimes wait to watch content until they are together. Searching for a video takes longer when co-watching because people must utilize other strategies or search terms than they would by themselves. The act of deciding on a video requires negotiating or turn-taking, which can lead to conflict at times. Finally, once a video is selected, people may continue to watch a video they do not enjoy because they do not know whether or not the other person is enjoying the video. This issue of pluralistic ignorance is exacerbated with computer-mediated communication where there are fewer visible cues from the other person. When collocated, people may be able to read from facial expressions or body movement whether someone else is enjoying a video that might be harder to glean from online communication [13]. Following, we describe the influence of impression management on the journey to co-watching along with suggestions to design remote and collocated co-watching experiences.

## Impression Management

In the same way that the act of video watching itself is a form of impression management [17], the video content chosen to co-watch may also contribute to a person's self-presentation. In Goffman's comparison of people's face-to-face interaction as a performance [10], he describes the “front stage,” the appearance a performer has when in front of the audience, and the “back stage,” when the performer is hidden from view. Watching videos privately could be seen as the “back stage” with co-watching as the “front stage.” Using this framework, it becomes clearer why finding a video takes longer and why other strategies need to be employed when co-watching.

Although participants reported commonly clicking on recommended content when alone, the videos that appear on a user's recommended videos may not be applicable to co-watching. Because these videos are selected by YouTube based on an individual's watch history when they are “back stage,” the videos can include content that is not interesting or relevant to all the people who are co-watching. Instead, people sharing a screen with others and logged into their YouTube account are akin to showing someone behind the scenes of a performance; even though it may be unintentional, the co-watcher is able to deduce the kind of content that the driver watches when alone.

As a result, people need alternate methods of finding content, such as using the search bar or relying on pre-watched content. By sharing pre-watched content, users are expecting the video to incite a certain response from the other person. Whether



or not the video is successful in this goal is a reflection of the user and his/her knowledge of the co-watcher. This match or mis-match can contribute to the impression that the viewer has of the sharer [15]. The concern for self-presentation does not end once a video has started; the driver of the device has the ability to return and select a new video. To determine if a new video should be selected requires correctly interpreting someone else's enjoyment of a video. Given the short duration of many YouTube videos, this process of picking a video that reflects positively upon oneself and determining whether or not to finish watching it, is a cycle that could repeat itself over and over again multiple times in a single viewing session.

Recent research in CSCW suggests that impression management of video content extends beyond YouTube. In their examination of MarathOn Multiscreen, a prototype that allowed collocated users to watch, share, and curate marathon footage, Anstead et al. reported that some users seized upon the opportunity to share a video to a TV whilst others were resistant to based on their embarrassment of the quality of the clip found [1]. The findings from this study are reflective of the same impression management that people engage in to choose YouTube videos for co-watching, and suggest that our findings might apply to other types of video content.

### Designing for Co-watching

Considering the entire co-watching journey reveals the opportunity to expand upon previous systems to explicitly support the steps that users take in order to co-watch videos.

Our results showed that co-watching can range in its intentionality and engagement with other co-watchers. Current popular video-platforms are situation agnostic, placing the burden on the user to adjust watching habits for different contexts. Since in the diary study our participants reported viewing in a variety of places, co-watching of short-form video presents a unique challenge that builds upon the difficulty of TV-selection for different audiences. One option would be for video players to take advantage of various sensors to better estimate when to present users with different content. For example, if combined with GPS, the video player could identify when users are in a public place and have content filtered for that type of viewing. The difficulties of sharing a device could also be alleviated by detecting connection to an external display through bluetooth or other sensors, and allow users to play a video on that display if in "sharing mode."

As revealed from the interviews, searching for video content when co-watching often takes longer than when alone, which could be addressed by designing features explicitly for groups. Having a group profile instead of an individual profile would curate recommended videos that are appropriate for co-watching; this would merge current "back stage" recommendations with ones that are appropriate for the "front stage." An alternate way of achieving appropriate recommendations would be for YouTube to merge different user's recommended videos based on common interests. The algorithm the system would use to aggregate votes for different videos could follow different strategies, as explored for group TV-watching by Masthoff [21]. While the model she proposed requires a

knowledge of explicit preferences and a stable group of watchers, which may not be the case with YouTube video-watching content, her results suggest that taking an average of people's preferences while also avoiding misery from any individual would be a potential strategy worthy of further exploration. By combining users' ratings of previous videos with the channels they follow, a similar model could be determined for YouTube content.

The issue of pluralistic ignorance observed in our study might not be as prevalent in TV watching since people might be more vocal about their opinions towards undesirable content if it's longer form, but classic TV literature suggests that pluralistic ignorance might also apply to TV co-watching [18]. "Family voting" which was previously proposed to balance out family-role dominance [28], could be applied to short-form video content as well. Designing the system to allow users to privately vote on videos being added to a shared queue could reduce pluralistic ignorance. This allows people to select a video in the "back stage" and only reveal it in the "front stage" when they are ready. Voting on which videos they want to watch could automatically rearrange the videos in the queue, allowing users to seamlessly incorporate each other's opinion in the shared experience without having to wait for verbal feedback. Voting in the group would allow each viewer's opinion to be acknowledged without privileging one user over another, minimizing the dominance that was found in our research to occur in certain relationships. This same interaction could be applied to collocated people broadcasting to a TV or another shared display. Using their phones, users could connect to the shared display to privately give their opinion without the potential of offending a co-watcher. While this follows an "everyone in control" scheme [23], allowing users to mediate the queue through their individual devices would alleviate complexities with sharing a singular control device.

Our findings also have applicability to the design of remote systems. Social TV could allow not only for audio during the video-watching stage, but also throughout the decision-making process [7]. Current remote systems propose entering a video URL to begin sharing [31], but such a system could be dominated by one user, leading to an uneven power balance. Instead, the system could explicitly support turn-taking. Once one user's video plays, the next user could be prompted to input the next video to be shared. This would follow the "one user at a time" scheme that was recommended by previous research [23]. Similarly, our research suggested that people may have different approaches to finding and assessing videos. A remote system could support not only selecting videos based on individual interests, but also browsing and searching behavior based on individual methods of looking for video content.

### CONCLUSION

This is one of the first studies to explore the short-form video co-watching journey: from searching for pre-watched content or taking the time to find new content, to negotiating or turn-taking to make a decision, to switching videos less when with another person. Through a survey, diary study, and reenactment sessions, we found that people engage each other at all

steps of the video process and use video sharing as a way of managing other's impressions of themselves. We presented key challenges in the co-watching journey and provided design suggestions to help to expand the social function of video sharing services. While we included an in-depth reflection on the methods above, we acknowledge that a task-based diary study is not without its limitations. Findings such as task completion times and difficulty ratings may have been impacted by the nature of the diary prompts. Similarly, assigning certain tasks to be completed alone and with another person could have resulted in the reporting of unnatural situations. Despite these limitations, tasks were successful in sparking discussion and reflection about their motivations for watching alone and co-watching. Future studies will look into whether the journey to co-watching applies to longer-form content, as well as assessing if our design recommendations are successful in more accurately reflecting the social engagement of watching videos together.

#### ACKNOWLEDGMENTS

We would like to thank Stephen Wheeler for help in creating the survey, Eileen Landay for recruiting participants for our studies, as well as all users who participated in the studies. We would also like to thank the reviewers for their helpful suggestions that improved upon previous versions of this paper.

#### REFERENCES

1. Edward Anstead, Steve Benford, and Robert Houghton. 2016. MarathOn multiscreen: group television watching and interaction in a viewing ecology. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 405–417.
2. Barry Brown, Moira McGregor, and Donald McMillan. 2014. 100 days of iphone use: Understanding the details of mobile device use. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services*. ACM, 223–232.
3. Marcus Carter, Bjorn Nansen, and Martin R Gibbs. 2014. Screen ecologies, multi-gaming and designing for different registers of engagement. In *Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play*. ACM, 37–46.
4. Allison JB Chaney, Mike Gartrell, Jake M Hofman, John Guiver, Noam Koenigstein, Pushmeet Kohli, and Ulrich Paquet. 2014. A large-scale exploration of group viewing patterns. In *Proceedings of the 2014 ACM international conference on Interactive experiences for TV and online video*. ACM, 31–38.
5. The Nielsen Company. March 2016. Video On Demand. (March 2016).  
<http://www.nielsen.com/content/dam/nielsen-global/eu/docs/pdf/Nielsen-global-video-on-demand.pdf>
6. Dan Cosley, Victoria Schwanda Sosik, Johnathon Schultz, S Tejaswi Peesapati, and Soyoung Lee. 2012. Experiences with designing tools for everyday reminiscing. *Human-Computer Interaction* 27, 1-2 (2012), 175–198.
7. Nicolas Ducheneaut, Robert J Moore, Lora Oehlberg, James D Thornton, and Eric Nickell. 2008. Social TV: Designing for distributed, sociable television viewing. *Intl. Journal of Human-Computer Interaction* 24, 2 (2008), 136–154.
8. Azadeh Forghani, Gina Venolia, and Kori Inkpen. 2014. Media2gether: Sharing Media during a Call. In *Proceedings of the 18th International Conference on Supporting Group Work*. ACM, 142–151.
9. 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 (CHI '11)*. ACM, New York, NY, USA, 311–314. DOI: <http://dx.doi.org/10.1145/1978942.1978986>
10. Erving Goffman and others. 1978. *The presentation of self in everyday life*. Harmondsworth.
11. Joseph K Goodman, Cynthia E Cryder, and Amar Cheema. 2013. Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples. *Journal of Behavioral Decision Making* 26, 3 (2013), 213–224.
12. Keith N Hampton and Neeti Gupta. 2008. Community and social interaction in the wireless city: wi-fi use in public and semi-public spaces. *New Media & Society* 10, 6 (2008), 831–850.
13. Sara Kiesler, Jane Siegel, and Timothy W McGuire. 1984. Social psychological aspects of computer-mediated communication. *American psychologist* 39, 10 (1984), 1123.
14. Tiffany CK Kwok, Michael Xuelin Huang, Wai Cheong Tam, and Grace Ngai. 2015. Emotar: Communicating Feelings through Video Sharing. In *Proceedings of the 20th International Conference on Intelligent User Interfaces*. ACM, 374–378.
15. Mark R Leary and Robin M Kowalski. 1990. Impression management: A literature review and two-component model. *Psychological bulletin* 107, 1 (1990), 34.
16. Benedikt Ley, Corinna Ogonowski, Jan Hess, Tim Reichling, Lin Wan, and Volker Wulf. 2014. Impacts of New Technologies on Media Usage and Social Behaviour in Domestic Environments. *Behav. Inf. Technol.* 33, 8 (Aug. 2014), 815–828. DOI: <http://dx.doi.org/10.1080/0144929X.2013.832383>
17. Trisha TC Lin, Jung Younbo, and Clarice Sim. 2015. Towards an understanding of intention to use mobile videos: Impression management, perceived facilitation, and social norms. *Mobile Media & Communication* 3, 1 (2015), 106–124.
18. James Lull. 1978. Choosing television programs by family vote. *Communication Quarterly* 26, 4 (1978), 53–57.
19. James Lull. 1982. How families select television programs: A mass-observational study. *Journal of Broadcasting & Electronic Media* 26, 4 (1982), 801–811.
20. Anna Macaranas, Gina Venolia, Kori Inkpen, and John Tang. 2013. Sharing Experiences over Video: watching video programs together at a distance. In

- Human-Computer Interaction–INTERACT 2013*. Springer, 73–90.
21. Judith Masthoff. 2004. Group modeling: Selecting a sequence of television items to suit a group of viewers. In *Personalized Digital Television*. Springer, 93–141.
  22. Tara Matthews, Kerwell Liao, Anna Turner, Marianne Berkovich, Robert Reeder, and Sunny Consolvo. 2016. She'll just grab any device that's closer: A Study of Everyday Device & Account Sharing in Households. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 5921–5932.
  23. Mark McGill, John Williamson, and Stephen A Brewster. 2014. How to lose friends & alienate people: sharing control of a single-user TV system. In *Proceedings of the 2014 ACM international conference on Interactive experiences for TV and online video*. ACM, 147–154.
  24. Kevin Mercer, Andrew May, and Val Mitchel. 2014. Designing for video: investigating the contextual cues within viewing situations. *Personal and ubiquitous computing* 18, 3 (2014), 723–735.
  25. Koji Miyauchi, Taro Sugahara, and Hiromi Oda. 2008. Relax or study?: A qualitative user study on the usage of Mobile TV and video. In *Changing Television Environments*. Springer, 128–132.
  26. Carman Neustaedter, Gina Venolia, Jason Procyk, and Daniel Hawkins. 2016. To Beam or Not to Beam: A Study of Remote Telepresence Attendance at an Academic Conference. (2016).
  27. 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 (CHI '07)*. ACM, New York, NY, USA, 857–866. DOI : <http://dx.doi.org/10.1145/1240624.1240754>
  28. Jisoo Park, Mark Blythe, Andrew Monk, and David Grayson. 2006. Sharable digital TV: relating ethnography to design through un-useless product suggestions. In *CHI'06 extended abstracts on Human factors in computing systems*. ACM, 1199–1204.
  29. Deborah A Prentice and Dale T Miller. 1996. Pluralistic ignorance and the perpetuation of social norms by unwitting actors. *Advances in experimental social psychology* 28 (1996), 161–210.
  30. Jenny Rosenberg and Nichole Egbert. 2011. Online impression management: personality traits and concerns for secondary goals as predictors of self-presentation tactics on Facebook. *Journal of Computer-Mediated Communication* 17, 1 (2011), 1–18.
  31. David A. Shamma, Marcello Bastea-Forte, Niels Joubert, and Yiming Liu. 2008. Enhancing Online Personal Connections Through the Synchronized Sharing of Online Video. In *CHI '08 Extended Abstracts on Human Factors in Computing Systems (CHI EA '08)*. ACM, New York, NY, USA, 2931–2936. DOI : <http://dx.doi.org/10.1145/1358628.1358786>
  32. Wei Song and Dian Tjondronegoro. A survey on usage of mobile video in Australia. In *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction*. 5–8.
  33. Jeroen Vanattenhoven and David Geerts. 2015. Contextual aspects of typical viewing situations: a new perspective for recommending television and video content. *Personal and Ubiquitous Computing* 19, 5-6 (2015), 761–779.
  34. Radu-Daniel Vatavu. 2015. Audience silhouettes: peripheral awareness of synchronous audience kinesics for social television. In *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video*. ACM, 13–22.
  35. Andrew M Webb, Chen Wang, Andruid Kerne, and Pablo Cesar. 2016. Distributed Liveness: Understanding How New Technologies Transform Performance Experiences. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 432–437.
  36. Bin Xu, Pamara Chang, Christopher L Welker, Natalya N Bazarova, and Dan Cosley. 2016. Automatic Archiving versus Default Deletion: What Snapchat Tells Us About Ephemerality in Design. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 1662–1675.