

Collaborative Online Video Watching

Justin D. Weisz

December 2009
CMU-CS-09-175

Computer Science Department
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Also appears as Human Computer Interaction Institute
Technical Report CMU-HCII-09-106

Thesis Committee:

Sara Kiesler (co-chair), Carnegie Mellon University
Hui Zhang (co-chair), Carnegie Mellon University
Luis von Ahn, Carnegie Mellon University
Wendy A. Kellogg, IBM T.J. Watson Research

*Submitted in partial fulfillment of the requirements
for the Degree of Doctor of Philosophy*

This work is licensed under a Creative Commons
Attribution-Noncommercial-Share Alike 3.0 United States License

<http://creativecommons.org/licenses/by-nc-sa/3.0/us/>

© 2009 Justin D. Weisz. *Some* rights reserved.

This research was sponsored by the National Science Foundation under grant numbers IIS-0325049, CNS-050187, CNS-0435382, and ANI-0331653. The views and conclusions contained in this document are those of the author and should not be interpreted as representing the official policies, either expressed or implied, of any sponsoring institution, the U.S. government or any other entity.

Keywords: Online video, collaboration, chat, social television, enjoyment, distraction, sociability, audiences, linguistic analysis, human computation, social proxies, MovieLens, YouTube, Facebook.

ABSTRACT

With the rise of broadband Internet, watching videos online has become a popular activity for millions of people. Many web sites encourage people to contribute, rate, and comment on video content, and more recently, to share their experiences with each other in real time. This dissertation explores the user experience of simultaneously chatting with other viewers while watching videos online.

Watching a video and talking with others is a form of multitasking that reduces information processing quality. The first part of this dissertation examines distraction, enjoyment, and sociability in collaborative watching. In a series of field studies and laboratory experiments, I show that small groups of friends and strangers enjoy chatting while watching videos together, despite chat's distractive effects in both text and audio channels.

The second part of this dissertation examines challenges of interacting with other viewers in large-scale broadcasts. I argue that viewers can chat with their friends and monitor the activities of the rest of the audience without feeling overwhelmed by using visual chat summaries and a social proxy representation of the audience.

The final part of this dissertation shows how three types of information can be inferred about a video from the raw chat log data of groups that watched the video together: a set of tags that describe the video, ratings of the video, and a profile of peoples' enjoyment of each part of the video. This information can be used to improve the quality of video search and recommendation engines, and provide behavioral-based feedback on viewers' enjoyment to content creators.

This dissertation provides new insight on the distraction from multitasking in an entertainment context. For the videos studied, it shows that although distraction does degrade information processing, it does not significantly harm the entertainment or social experience. This dissertation also provides concrete designs and recommendations for user interfaces for large-scale online video broadcasts. Finally, this dissertation demonstrates that information about a video that would be difficult or impossible to infer through a computer algorithm can be learned from the social interactions that occur as viewers watch together.

Dedicated to the four I lost

my sister, Pamela Brooke Weisz

my dog, Heidi Rachel Weisz

my grandfather, Bernard Filman

my cousin, Jenna Filman

Baruch dayan ha'emet

ACKNOWLEDGEMENTS

Bridges are an important part of Pittsburgh culture. Our city has over 440 bridges; more than Venice! My thesis represents a bridge between human-computer interaction and computer networking by demonstrating how the methods and design principles of the former can be used to create and evaluate novel applications for the systems of the latter. My sincerest thanks goes to my advisors and committee for providing me the opportunity and means to create this bridge.

Sara Kiesler guided me through the process of transforming vague ideas about online video into a full-fledged research agenda, and helped me carry out that agenda. Hui Zhang taught me the zen of thinking big and being realistic at the same time. Luis von Ahn reminded me that practical applications are highly important, and especially so when studying entertainment experiences. Wendy Kellogg taught me that quantitative and qualitative methods are both important for reaching understanding. Tom Erickson, extraordinarily generous in his role as an unofficial committee member, reassures me that I actually have talent as a designer (although I am still skeptical of this). I could not have asked for a better thesis committee.

I also could not have asked for a better institution in which to conduct my research. Carnegie Mellon has provided me with ten of the most demanding, most enjoyable, and most rewarding years of my life. I am ever thankful to whomever called me in 1999 to ask if I wanted to attend the undergraduate program in computer science. My answer remains, “absolutely!”

During my graduate studies, I have had the pleasure of working with many talented, inspirational, and creative undergraduate and graduate students, including Katie Appleton, Gem Bleasdell, Caitlinn Cork, Sean Curtis, Andy Echenique, Brian Ellis, Allison Gallant, Ming Guo, Wenyao Ho, Holger Kuehnle, Lily Li, Lisa Ly, Kimmy Nederlof, Derek Parham, Tamir Sen, Afeique Shekh, Rushikesh Sheth, Hao Su, Jiwoo Suh, Margaret Szeto, Harshad Telang, Jessica Wu, and Bei Yang.

I offer thanks to many of the students, faculty, staff, mentors, and friends who have provided me with wisdom, criticism, motivation, support, and tremendously fun dinners throughout the years, including Bryce Aisaka, Turadg Aleahmad, Shazad Ali,

Caleb Astey, Daniel Avrahami, Aruna Balakrishnan, Patrick Barry, Karl Becker, Rachel Bellamy, Katie Bessiere, Ashwin Bharambe, Daniel Bilenko, Alex Billington, Devin Blais, Art Boni, Michael Bridges, Matt Brown, Moira Burke, Deborah Cavlovich, David Casillas, Pablo Cesar, Darsen Chen, Jim Christensen, William Courtright, Daniel Cox, Liz Crawford, Laura Dabbish, Catalina Danis, Scott Davidoff, Uri Dekel, Sara Drenner, Andrew Dubois, Jennifer Dubois, Matt Easterday, Khalid El-Arini, Jason Ellis, Dan Frankowski, Matt Fredrikson, Sandeep Gaan, Aditya Ganjam, David Geerts, Daniel Golovin, Pat Gunn, Christine Halverson, Daniel Hamad, Gunnar Harboe, Mor Harchol-Balter, Max Harper, Laura Hiatt, Ross Higashi, Paul Homlish, Jason Hong, Garrison Hu, Travis Iwanaga, Farokh Jamalyaria, Dominic Jonak, Yan Ke, Greg Kesden, Kevin Killourhy, Aniket Kittur, Andrea Knight, Andy Ko, Joseph Konstan, Adam Kramer, Robert Kraut, Brian Krausz, Neel Krishnaswami, Lewis Kwon, Erren Lester, Curt Lewis, Ian Li, Julio López, Bruce Maggs, Peter Malkin, Pratyusa Manadhata, Joey Markowicz, Kathy McNiff, Lauren Merski, Neema Moraveji, David Murray, Andy Myers, Mukesh Nathan, Robert Olczak, Reid Priedhorsky, Sandeep Panday, Jeffrey Pang, Brian Pao, Kayur Patel, Chytra Pawashe, Michael Piatek, Pablo Quinones, Martina Rau, Yuqing Ren, Greg Reshko, John T. Richards, Steve Riggins, Jennifer Rode, Carolyn Rosé, Joseph Round, Deanna Rubin, Jeffrey Sander, Mary Scott, Srinu Seshan, Peter Scupelli, Vyas Sekar, Shilad Sen, Jahanzeb Sherwani, Michael Shigemoto, Mark Stehlik, Jeffrey Stylos, Jeremy Sussman, John C. Thomas, Niraj Tolia, Cristen Torrey, Arthur Tu, Miki Urisaka, Tracee Vetting-Wolf, JD Vogt, Jason Weill, Adam Wolbach, Jeff Wong, Lena Yoo, Shawn Yoon, Jibin Zhan, Noam Zeilberger, all of the OGS, Karate, and Hawaii guys, and many others.

My sister Pamela deserves special acknowledgement. My motivation, my temperament, and my ambition to make life more enjoyable for others are all direct products of her influence. Without her, the entire course of my life would have been different. Without her, this dissertation would never have been written. Pammy's life is a testament to the fact that all life is sacred and all lives have meaning.

My parents Shari and Sandy, and my little sister Nicki, have my most sincere thanks and deepest gratitude for being the most incredible people I will ever know. We have learned together that life is cruel, unfair, and unforgiving. We have also learned that it is wonderful and awesome, and a gift to be treasured and enjoyed every single day whilst we are able. I thank you for the opportunity and for sharing it with me.

My final thanks go to Jim Forde and Alan Forsythe, whom I have not forgotten.

TABLE OF CONTENTS

Abstract	iii
Acknowledgements	vii
Table of Contents	ix
List of Figures	xv
List of Tables	xix
1. Introduction	1
1.1. The Rise of Video on the Internet	2
1.2. Television, Online Video, and Social Capital	5
1.3. Human Factors and Human Attention	8
1.4. Thesis Statement	10
1.5. Contributions and Significance.....	11
1.6. Overview	13
PART I: BACKGROUND	
2. Combining Chat With Video: Related Work	19
2.1. Television.....	19
2.2. Computer-Mediated Communication and Media Richness.....	21
2.3. Applications of Chat with Video.....	24
2.4. Summary and Conclusions	28
3. A Framework for Collaborative Online Video	31
3.1. Content.....	32
3.2. Distribution Technology.....	33
3.3. Viewing Device	36
3.4. Interaction	37
3.5. Playback Model.....	40
3.6. Summary and Conclusions	41
4. Social Online Video: A Survey of Watching Videos on YouTube	43
4.1. A Survey of YouTube Users.....	44
4.2. Basic Usage Patterns	46
4.3. Value and Benefits.....	47

4.4.	Finding and Sharing Videos	49
4.5.	Social Behaviors	53
4.6.	Personality	54
4.7.	Summary and Conclusions	61

PART II: SIMULTANEOUS WATCHING AND CHATTING

5.	Empirical Studies of Chatting While Watching	65
5.1.	General Methods	67
5.2.	Measures	68
5.3.	Videos	70
5.4.	Summary and Conclusions	74
6.	The MovieLens Study	77
6.1.	Research Questions.....	78
6.2.	Design	78
6.3.	Participants	78
6.4.	Method	78
6.5.	Measures	80
6.6.	Results	80
6.7.	Discussion	82
6.8.	Summary and Conclusions	83
7.	The Chat Distraction Study	85
7.1.	Simplifying Text Chat	85
7.2.	Research Questions.....	87
7.3.	Design	87
7.4.	Participants	89
7.5.	Method	89
7.6.	Measures	89
7.7.	Results	90
7.8.	Discussion	91
7.9.	Summary and Conclusions	92
8.	The Cartoon Study	93
8.1.	Research Questions.....	95
8.2.	Design	96
8.3.	Participants	96
8.4.	Method	96
8.5.	Measures	98

8.6.	Results	99
8.7.	Discussion.....	108
8.8.	Summary and Conclusions	110
9.	The Text vs. Audio Study	113
9.1.	Research Questions	116
9.2.	Design.....	117
9.3.	Participants	117
9.4.	Method.....	118
9.5.	Measures.....	120
9.6.	Results	122
9.7.	Discussion.....	131
9.8.	Summary and Conclusions	135
10.	Overall Discussion of the Empirical Studies	137
10.1.	Sociability	137
10.2.	Distraction.....	139
10.3.	Design Recommendations.....	141
10.4.	Summary and Conclusions	142
PART III: LARGE-SCALE COLLABORATIVE WATCHING		
11.	Designing for Large Virtual Audiences	145
11.1.	The Challenges of a Large Audience.....	146
11.2.	Finding Chat Groups	150
11.3.	Maintaining Awareness of the Audience.....	155
11.4.	A Social Proxy for Large Audiences.....	162
11.5.	Evaluation of Audience Representation and Chat Summarization..	167
11.6.	General Discussion	178
11.7.	Summary and Conclusions	181
12.	The “Social Video” Application	183
12.1.	Scenarios	183
12.2.	Design Goals	184
12.3.	Design Decisions	186
12.4.	Implementation	187
12.5.	General Usage	196
12.6.	Evaluation Study	197
12.7.	Interface Redesign.....	204
12.8.	General Discussion	208

12.9.	Summary and Conclusions	209
-------	-------------------------------	-----

PART IV: LEARNING ABOUT VIDEOS

13.	Learning About Videos From Chat Data	213
13.1.	Online Social Interactions as Human Computations	215
13.2.	Collecting and Mining Chat Transcripts	216
13.3.	Summary and Conclusions	218
14.	Tag Extraction	219
14.1.	General Method for Tag Extraction	220
14.2.	Tag Evaluation Study	223
14.3.	General Discussion	233
14.4.	Summary and Conclusions	235
15.	Inferring Video Ratings	237
15.1.	Inferring Ratings From Linguistic Features	238
15.2.	Data Set	238
15.3.	Regression Models.....	239
15.4.	Discussion.....	243
15.5.	Summary and Conclusions	245
16.	Learning Enjoyment Profiles	247
16.1.	Prior Work.....	247
16.2.	Aggregating Laughter	249
16.3.	Extracting Enjoyment Profiles From Chat Data	250
16.4.	Enjoyment Profile Accuracy Study	254
16.5.	General Discussion	265
16.6.	Summary and Conclusions	269

PART V: CONCLUSIONS

17.	Limitations and Future Work	273
17.1.	Alternative Video Content	273
17.2.	Collect More Real-World Data.....	275
17.3.	Additional Behavioral Measures	276
17.4.	Visualization Distraction.....	277
17.5.	Social Dashboards.....	277
17.6.	Improved Text Mining.....	278
17.7.	Summary and Conclusions	279

18. Conclusions	281
Appendix A: Popular Online Video Sites and Systems	283
Appendix B: Scales & Measures	287
Appendix C: Laughter Extension to LIWC	293
References	295

LIST OF FIGURES

Figure 1-1. The inauguration of President Barack Obama in 2009, live online with CNN and Facebook.	1
Figure 1-2. Timeline of the studies and design projects presented in this dissertation and their corresponding chapters. Arrows represent the approximate flow of ideas and influence between studies as well as data when noted. The height of the boxes is generally not reflective of the amount of time spent on each project. Boxes in the outline style represent projects focused specifically on design and implementation.	15
Figure 3-1. Five aspects of the collaborative online video experience.....	32
Figure 6-1. Screenshot of the ESM software (minimized) playing a movie in Windows Media Player. The text chat feature (shown) enabled simultaneous viewers to chat with each other while watching.....	79
Figure 7-1. Screenshot from UStream.TV showing video and text chat.....	86
Figure 7-2. Screenshots from the Chat Distraction study. (a) Full chat history condition. (b) No chat history condition. (c) No chat condition.....	88
Figure 8-1. Different arrangements of video and chat windows. Participants were allowed to move and resize the windows to their individual preference. (a) Default arrangement with chat offset from video. (b) Full-screen video with chat overlaid. (c) Chat and video side-by-side.....	97
Figure 8-2. Distribution of cartoon ratings for groups with and without chat, by cartoon quality. Error bars represent 95% confidence intervals. The difference in mean rating for poor cartoons was significant ($p = .02$); the differences in mean rating for okay and good cartoons are not significant.	100
Figure 8-3. Effect of chat on liking of others, between groups of friends and groups of strangers, with and without chat. Error bars represent 95% confidence intervals.....	106
Figure 8-4. Effect of chat on feelings of closeness, between groups of friends and groups of strangers, with and without chat. Error bars represent 95% confidence intervals.	107
Figure 9-1. Virtual avatars watch a movie together using the Netflix application on Xbox 360. The real viewers to whom those avatars belong communicate with each other using voice chat.....	116
Figure 9-2. Screenshot from the Text vs. Audio study. The text chat feature was not displayed for participants in the no-chat and voice-chat conditions. Participants with voice chat or both text and voice chat could speak to each other over their headsets.....	119
Figure 9-3. (a) Participants watch the same videos at the same time. Each participant's video is synchronized with the other members of their group. (b) Participants watch the videos in a different, random order from each other. The	

initial video for each participant is always different. Transitions between successive videos do not occur at the same points in time as the videos are of different lengths.....	120
Figure 9-4. Preferences for different chat media by chat media condition. Each horizontal strip shows the distribution of media preferences for the participants in the specified condition.....	124
Figure 11-1. The Facebook Live Stream Box. This widget can be coupled with any live video stream online to create a branded, collaborative video experience for viewers with Facebook accounts. Names and photos have been blurred to preserve anonymity. Because viewers may not see all of the messages sent, back-and-forth conversation may be difficult or impossible. When messages are filtered out, it is unclear to message senders who will see their message. Further, for viewers who reply to a message, it is unclear if the original sender will see the reply.....	147
Figure 11-2. Tag cloud of my thesis proposal, created by wordle.net.....	160
Figure 11-3. Twistori visualization of messages on Twitter. Messages are selected on the basis of containing phrases such as “I feel” or “I think.” New messages are added at a rate of about one per second.....	161
Figure 11-4. We Feel Fine visualization of feelings and sentiments made in blog posts. Individual messages are represented with small icons that swarm around the interface. Messages can be viewed by clicking on them. Color represents the feeling described by the message.....	161
Figure 11-5. General concept for the large audience proxy. The audience is presented from the perspective of an individual user. Two dimensions are used to place audience members: angular position and distance from the center.....	163
Figure 11-6. Discretization of angular position and distance to center for nominal or ordinal attributes. (a) Four angular categories and two distance-based categories of equal size. (b) Four angular categories and three distance-based categories with areas proportional to the number of audience members in each category.	164
Figure 11-7. Using color to encode attributes of the user. (a) A color highlight around the user’s icon can represent an attribute encoded by distance to the center. (b) A line can be used to represent a continuous attribute represented by angular position. (c) A border, edge, or background highlight can be used to represent an ordinal or nominal attribute of the user.	164
Figure 11-8. Prototype of the audience proxy for watching online video in a large audience. Chattiness is a continuous measure encoded in angular position. Relationship is a discretized measure (friend vs. stranger) encoded in distance to center. The viewer and his friends are displayed using their pictures to convey a greater degree of presence and connection; strangers are displayed using generic icons. Friends are also placed “in” the current group or “out” in other chat groups.....	167
Figure 11-9. Screenshots from the prototype interface for watching a live video event in a large audience. Names have been blurred to preserve anonymity. Top: The audience size display (left) shows only the size of the audience and the number of friends in the audience; the tag cloud (right) shows a summary of the popular topics of chat. Bottom: The audience proxy (left) shows the viewer, their friends in the audience, and groups of other audience members; the scrolling list (right) shows chat messages from viewers in other chat groups.	168

Figure 12-1. The Social Video application. This interface was situated on the Facebook canvas page. Advertisements and Facebook toolbars have been removed. Names have been blurred for anonymity. (a) Visualizations of system activity. (b) List of online friends. (c) Multiple persistent chat groups can be joined. (d) Unread message counts allow users to monitor activity across their chat groups. (e) Status messages in chat display users' playback activity. (f) Group ratings help people find interesting groups. (g) Video playback status is visible for all users.....	188
Figure 12-2. The shared video playlist. The "Start Next Video" button caused viewers to immediately start playing the next video in the playlist.	189
Figure 12-3. The waiting screen. This screen was shown to viewers once the current video finished playing for them. The playback status of each person in the group was shown. Names have been blurred to preserve anonymity.	191
Figure 12-4. Latest chat visualization. Chat messages were randomly selected from public chat groups, with a preference for newer messages. Names have been blurred to protect anonymity.	195
Figure 12-5. Number of registered users over time. Overall, 247 users signed up to use the application. The highlighted period from May 13 to May 26 corresponds to the evaluation study discussed in Section 12.6.	197
Figure 12-6. Patterns of watching and chatting in the HIT sessions. Colored bands indicate video, gaps between bands indicate waiting, and black dots indicate chat. Only the first 80 minutes of activity are shown.	201
Figure 12-7. Amount of chat while watching the videos, paused during a video, or waiting to begin the next video.	203
Figure 12-8. (top) Mockup of the redesigned Social Video interface. Tabs are used on the bottom to provide access to friends, videos, and visualizations of other users' activity. The blue boxes on the right are a placeholder for a list of people in the current chat group. (bottom) Screenshot from the current implementation of the redesigned Social Video interface. Names have been blurred for anonymity.	206
Figure 12-9. Prototypes of the (a) groups, (b) friends, and (c) videos tabs. These tabs are displayed below the video and can be used without interrupting video playback or triggering a page refresh (unlike the initial interface).	207
Figure 13-1. Algorithmic process for collaborative online video watching as a human computation. (a) Inputs to the algorithm are groups of people and a video. (b) The social activity of watching videos and chatting is used to produce a chat transcript for each group. A chat filter can be used to convert voice chat into a text transcript, and text transcripts can be normalized before analysis. (c) Chat transcripts are a by-product of the social experience. (d) Various analyses are performed on the transcripts to produce the outputs of tags, ratings, and enjoyment profiles.....	217
Figure 14-1. Process diagram for extracting tags for a video from chat transcripts. (a) Multiple chat transcripts for a video are combined into a single transcript. (b) Individual comments are normalized. (c) Unigrams and bigrams are extracted from the comment set. (d) Each term is weighted; in this example, the TF-IDF metric is used to weight terms. (e) The top K terms are chosen as the extracted tags.....	220

Figure 14-2. Formulas for computing TF-IDF weights. TF (term frequency) is the frequency of occurrence of term t during video v . IDF (inverse document frequency) is the measure of a term’s “uniqueness,” and is inversely proportional to the number of videos in which that term was used.....221

Figure 14-3. Comparison of tag relevancy ratings between native and non-native English speakers. Error bars represent 95% confidence intervals. All differences between native and non-native speakers were significant at the $p = .05$ level, except for Common tags.....229

Figure 14-4. Mean per-tag relevance for each tag source and each video. Student’s t letters are listed in parentheses and show which tag sources were significantly different at the $p = .05$ level. Tag sources not connected by the same letter were significantly different.....230

Figure 16-1. Hypothetical enjoyment profile. Shaded regions indicate the locations of the high points.....250

Figure 16-2. Enjoyment profiles for several videos in the Cartoon and Text vs. Audio studies. The number of utterances of laughter is listed in parentheses. Time is represented on the x-axis in seconds. The relative amount of enjoyment (measured by amount of laughter) is shown on the y-axis.....254

Figure 16-3. (a) Comparison of the hand-created profile with the chat-extracted profile for “Ali G - War”. (b) Absolute error between the profiles.....257

Figure 16-4. Equations for computing percentage agreement (A) between chat-extracted and hand-created profiles.....258

Figure 16-5. Enjoyment profiles for (a) “Daughters” and (b) “Fuggy Fuggy.” The blue line (darker) is the hand-created profile; the yellow line (lighter) is the chat-extracted profile.....260

Figure 16-6. Hypothetical enjoyment following a bimodal distribution. This example illustrates why standard measures of spread and peakedness cannot determine consensus in enjoyment profiles. In this case, all viewers have labelled the beginning and end parts of the video as their favorite part.....261

LIST OF TABLES

Table 1-1. Several of the online video sites and systems discussed in this dissertation. A more complete version of this table is given in Appendix A.....	4
Table 1-2. The studies discussed in this dissertation.....	15
Table 2-1. Five core problems with text chat, adopted from Smith, Cadiz, and Burkhalter (2000).....	22
Table 3-1. Methods for online video delivery.....	35
Table 3-2. Time-space taxonomy applied to online video interactions.....	38
Table 3-3. Comparison between the playlist and streaming video models.....	41
Table 4-1. Summary of research questions for the survey of YouTube users.....	44
Table 4-2. Demographics of survey respondents and the general student population. (*) Faculty and staff responses have been excluded to compare with enrollment data.....	45
Table 4-3. YouTube enjoyment scale. (*) Item 2 was dropped in the analysis to increase reliability.....	48
Table 4-4. Reasons for using YouTube. Percentages of respondents who affirmed each reason are listed in parentheses.....	49
Table 4-5. Frequency of usage for different types of video recommendations. "Frequent" is the sum of "Often" and "Almost always" responses. Reported numbers are the percentage of respondents in each category.....	50
Table 4-6. People with whom videos are shared. "Frequent" is the sum of "Often" and "Almost always" responses. Reported numbers are the percentage of respondents in each category.....	51
Table 4-7. Methods used for sharing videos. "Frequent" is the sum of "Often" and "Almost always" responses. Reported numbers are the percentage of respondents in each category.....	51
Table 4-8. Motivations for uploading videos to YouTube. Percentages are of N = 37 people who reported uploading videos.....	52
Table 4-9. Activities performed while watching videos. "Frequent" is the sum of "Often" and "Almost always" responses. Reported numbers are the percentage of respondents in each category.....	53
Table 4-10. Distributions of personality constructs. Responses are on a 5-point scale, with 5 representing the largest value for each construct.....	54
Table 5-1. Summary of the empirical studies of collaborative online video watching.....	66
Table 5-2. List of videos used in each study. Cartoon study videos came from Channel Frederator, a cartoon video podcast. Text vs. Audio (and Chat Distraction) study videos came from YouTube. The runtime of the video and a brief description of the video's genre and content are given.....	71

Table 5-3. Ratings for each video across the empirical studies. All ratings are on a 5-point scale, with 5 as the highest rating. Standard deviations are listed in parentheses. Missing values indicate that a video was not used in the corresponding study. For the Text vs. Audio pretest, the number of raters for each video is listed in square brackets.	73
Table 6-1. Distribution of chat topics across the movie showings.....	82
Table 8-1. Distribution of chat categories. Examples of chat are printed in their original form. Lines of chat were coded into only one category, except for laughter. Lines of chat containing laughter were coded as either solely consisting of laughter (7.4%) or containing laughter in addition to other content (9.4%).	104
Table 9-1. Distribution of chat content across coding categories. The overall distribution is given, as well as distributions for groups based on video order and chat media.	130
Table 11-1. Popular television and online video events that have attracted audiences of millions. The source of the estimated viewership is listed, along with the date the event occurred.....	146
Table 11-2. Dimensions of chat groups that can be used as a basis for recommending those groups to users.....	151
Table 11-3. Self-reported preferences for different features of chat groups. Weights are the sum of the weights assigned to each category across all respondents. Respondents could express interest for multiple types of groups in each category except for (*) social network. (**) This answer choice was not present on the survey and was written in by two respondents.....	152
Table 11-4. Visualizations that provide awareness of users and their status, activities, and/or interactions with each other. The design elements used for representing users, their status, and their activities in the system are given.....	156
Table 11-5. Visualizations that summarize the contents of users' interactions, such as their chat or status messages.	159
Table 11-6. Reported cutoffs for different audience sizes. (*) This participant referenced the population of China when thinking about what constituted a "large" audience.	172
Table 12-1. Videos used in the evaluation study. The number of videos in each set and the number of HIT groups that watched those videos are given.	199
Table 12-2. Summary of watching and chatting patterns.....	202
Table 12-3. Social Video design goals and how they are achieved in the redesigned user interface.	207
Table 14-1. Tag sets for each video in the Tag Evaluation study. Common tags marked with an asterisk (*) also belonged to the set of tags extracted from chat. Common tags marked with a dagger (†) also belonged to the set of tags from YouTube. All common tags also belonged to the set of tags from human raters. (‡) This tag was unintentionally relevant to the video and thus was excluded from the analysis.....	225
Table 14-2. Relevance statistics for tag sources. Standard deviations are listed in parentheses. Student's t letters show which tag sources were significantly different at the p = .05 level. Tag sources not connected by the same letter were significantly different.....	228

Table 14-3. Percentage of tags in the Common set that came from chat extraction and YouTube.....	231
Table 14-4. Relevancies for tag sources, without excluding Common tags. Change scores show the difference in mean (SD) per-tag relevance from their counterparts in Table 14-2.....	232
Table 15-1. Format of the data set for predicting video ratings from linguistic features. The values of each LIWC category were a count of the number of words typed that matched that category. Ratings ranged from 1 (low) to 5 (high). Participants had one row in this data set for each video they watched.	239
Table 15-2. Descriptive statistics of per-participant linguistic features and regression models for predicting video ratings from these features. Mean values are of the number of words spoken in each category. Unstandardized regression coefficients are reported for each model. Standard errors (SE) on regression coefficients are given in parenthesis. Level of significance for the coefficients are reported as (*) $p < .10$ and (**) $p < .05$. The table data pertains to 88 participants with only text chat in the Cartoon and Text vs. Audio studies.....	240
Table 16-1. Amount of laughter for the videos across all groups in the high and low data conditions. Standard deviations are listed in parentheses.	252
Table 16-2. Several categories of reasons for why participants labeled specific parts as being their favorite parts. Commenters are labelled with their study ID, and comments are reproduced in their original form.....	256
Table 16-3. Percentage agreement and correlations of enjoyment mass over time for both 5-second and 15-second resolutions. (*) $p < .05$ for the correlations.	258
Table 16-4. Results of clustering participants to determine consensus on labeling the best parts of videos. Only participants that reported at least one favorite part for a video were included in the clustering; this count is reported as the rater count. Video ratings were made on a 5-point scale (5 highest). The largest cluster ratio is the ratio of the largest cluster size to the rater count.	262
Table A-1. List of popular online video sites and systems. Sites were selected for inclusion on the basis of popularity, uniqueness of content, social interaction features, or discussion in this dissertation. The year listed is the year in which the site was founded, the video component of the site was launched, or the technology was released. Descriptions of each category are given and are generally applicable to each site listed in that category; additional description is given for each site where appropriate.....	283
Table C-1. Regular expressions for classifying textual laughter.....	293

1.

INTRODUCTION



Figure 1-1. The inauguration of President Barack Obama in 2009, live online with CNN and Facebook.

On January 20th, 2009, President Elect Barack Obama was sworn in as the 44th President of the United States of America. Like prior ceremonies, millions of television viewers watched this process unfold live, as it occurred. Unlike past ceremonies, this event was the single most-watched event in the history of online video to date (Sutter, 2009). In addition, it was the first oath of office to combine the television production of CNN with the social networking reach of Facebook to provide 7.7 million viewers an opportunity to share the moment by expressing their

thoughts and feelings with each other. This event heralded the beginning of a new era – that of collaborative online video watching.

1.1. THE RISE OF VIDEO ON THE INTERNET

Video on the Internet has been available for over a decade. Companies including RealNetworks, Sorenson, Apple, and Microsoft developed the early technologies that made streaming video online possible. However, mainstream availability and consumption of online video did not come to pass until two important enabling preconditions were met.

The first enabler of mainstream online video was the increase in penetration of home broadband Internet connections. Broadband technologies such as cable models and digital subscriber lines (DSL) provide users with an order of magnitude more bandwidth than the analog modems of the prior decade. The transition from kilobits per second to megabits per second allowed for the rapid delivery of high-fidelity content over the Internet, including high-quality streaming video. According to a Pew Internet and American Life study by Horrigan and Smith (2007), the percentage of Americans with broadband Internet connections surpassed the percentage of Americans with dial-up Internet connections around February, 2005. In fact, this trend toward higher-bandwidth Internet connections is occurring on a world-wide scale. As of this writing¹, of the 20 countries with the highest number of broadband subscribers, 13 have penetration rates above 15%.

Widespread availability of video content is the second enabler of mainstream online video; without content available to watch, there would be no online video of which to speak! In 2005, a small startup company was founded to capitalize on the increasing availability and popularity of home broadband Internet access. Its goal was to provide fast and easy access to online videos, and allow people to share videos with each other. This company was YouTube, and four years later, it stands as the most popular online video site, attracting hundreds of millions of visitors every day (comScore, 2007 & 2009).

¹ Data retrieved from Internet World Stats on October 15, 2009. <http://www.internetworldstats.com/dsl.htm>. The penetration rate is computed as the ratio of the number of broadband subscribers to the population of the country.

The satisfaction of these preconditions has, in fact, transformed watching videos into a mainstream activity online. Another Pew study by Madden (2007) finds that 57% of adult Internet users surveyed reported watching or downloading video content online. For young adults, (ages 18-29) consumption of online video is higher, with 76% reporting that they watch videos online.

YouTube ushered in a golden era of online video, and its format – allowing users to upload, share, rate, and comment on video clips – has been imitated by countless sites. Many of these sites are focused on providing entertainment experiences by allowing viewers to watch and share content produced by their members (amateur content) or by professionals. For example, three of the major broadcasting networks in the United States – NBC, ABC and CBS – allow viewers to watch their favorite TV shows online.

Online video has extended beyond keeping viewers entertained. Sites like TED.com and MIT's OpenCourseWare specialize in educational material by making lectures and talks available to the public. Apple's podcast directory (accessible from within iTunes) features many podcasts focused on education and language learning. Political events and the dissemination of news and current events are also turning to online video to engage their audiences, promote discussion, and raise awareness of important issues. Sites including Current.TV and LiveLeak specialize in newsworthy content, and shows such as Comedy Central's Daily Show and PBS's Bill Moyers Journal are readily available online. Sports are also popular, and sites like JumpTV allow viewers to watch sporting events from all around the world. Religion has also found an audience through online video, with many priests, pastors, rabbis, and congregants broadcasting their lectures, sermons, and stories on sites such as GodTube and JewTube. Finally, I would be remiss not to at least mention pornography, which has a long history of creating a demand for new technologies such as online video, web cams, video conferencing, secure credit card transactions, and live interaction online.

All of the sites mentioned above demonstrate that online video is becoming increasingly pervasive in our culture, and that there are many options available to Internet users to consume video online. To highlight these options, I summarize some of the currently popular online video sites in Table 1-1. For historical value, a more thorough list of online video sites is given in Table A-1 in Appendix A. These sites are differentiated in terms of the content they provide, the technology they use

to distribute video, or the features they provide for enabling social interaction among viewers. In Chapter 3, I present a framework that more thoroughly describes these different aspects of online video.

Table 1-1. Several of the online video sites and systems discussed in this dissertation. A more complete version of this table is given in Appendix A.

Category	Description	Examples
User generated content – upload	Users upload home videos; content publishers upload television & movie clips, music videos, etc.	YouTube, Yahoo Video
User generated content – streaming	Stream live video from computers, game consoles, mobile devices, etc.	Justin.TV, UStream.TV
Education / Information	Video sites providing educational materials and sharing inspired thinking	TED, OpenCourseWare
News, politics, & current events	Sites and shows focused on keeping viewers informed about news, politics, and current events	C-SPAN, CNN/ Facebook, Current.TV, LiveLeak
Television & movies	Major networks & studios provide access to their television and movie content online	NBC, ABC, CBS, Hulu
Sports	Sports content; live sporting events	JumpTV, ESPN
Pornography	Live webcams with chat; user generated content	YouPorn, RedTube
Social TV research systems	Systems developed specially for research in social and interactive television	AmigoTV, Social TV, Social Video, Zync
Peer-to-peer systems (P2P)	Applications that allow users to publish and view live streaming video over the Internet using peer-to-peer technologies	ESM, PPLive, Sopcast

As can be seen from Tables 1-1 and A-1, “online video” encompasses a very broad spectrum – there are many types of content available through many different web sites and applications. One common characteristic shared by each of these types of online video sites and applications is that they possess the opportunity to enable and foster social interaction among their users. Shared consumption of video media, such as TV and movies, enables groups of viewers to interact with each other, during the act of consumption, to create a new experience. Fans of popular sporting events understand how transformative the social aspect is to the act of consumption. Watching a big game with others – friends or strangers, in public or in private – is generally preferable to watching it alone (Vosgerau, Wertenbroch, & Carmon, 2006).

Despite the ability of television to foster social experiences, in many cases, television is experienced in isolation. People who are alone often greatly enjoy television, movies, and music, and use these media as an escape from their everyday cares

(Finn & Gorr, 2001; Hills & Argyle, 1998). For those viewers who want a social viewing experience, television requires viewers to be physically co-located. This requirement restricts potential social interactions to only those who are present together.

Communications technologies remove boundaries to interaction. Technologies like the telephone and the Internet allow remote viewers to communicate with each other while watching together. Thus, they enable viewers to have a social viewing experience by providing a communications channel to remote partners. This dissertation examines the ability for remote viewers to have a sociable and enjoyable experience while watching video content together online.

The next sections discuss two perspectives on the desirability of collaborative viewing: that it is desirable because of the potential benefits to social capital, and that it is undesirable because multitasking between watching videos and chatting is distracting and frustrating, which may lead to a poor experience.

1.2. TELEVISION, ONLINE VIDEO, AND SOCIAL CAPITAL

Social capital is the notion that relationships between people have a productive quality to them. By having strong relationships with others, we increase our opportunities for receiving support when needed, such as social, emotional, financial, informational, and health support. In addition, by having weaker ties to many others, we increase our ability to find resources we need within our social networks, such as people with a particular expertise or knowledge. Social capital is associated with positive individual and collective outcomes, such as better health (Lochner et al., 1999; Parker et al., 2001), better education (Putnam, 2001, Chapter 17), economic development (Putnam, 2001, Chapter 19), and good government (Putnam, 1993).

Robert Putnam, a preeminent researcher of social capital, has argued that social capital has been on the decline in America for the past several decades (Putnam, 2001). The driving forces behind this decline are decreased participation in social organizations such as team sports and civic, volunteer, and religious organizations. In his argument, Putnam specifically implicates television-watching as one of the causes of this decrease in participation (Putnam, 1995). Instead of participating in a bowling league or the PTA after work – places where people can form new

relationships and expand their social networks – people instead go home to watch the latest sitcom or reality TV show. Although it has been argued by Norris (1996) that news and current affairs programs “[do] not seem to be damaging to the democratic health of society” (Norris, 1996, p. 479), the act of watching television is unique among other forms of media consumption because its consumption inhibits participation in other activities (Putnam, 1995). Thus, watching television in the home precludes participation in activities outside of the home, reducing one’s opportunities to grow and strengthen their social networks.

Online video provides opportunities for viewers to interact with each other around television content. Because the Internet is coupled with the act of media consumption, technologies that allow viewers to communicate, interact, and share with each other can be leveraged to create a new, *social* viewing experience. In essence, these technologies can be used to transform the traditionally “lean-back” experience of watching TV into a “lean-forward” experience of watching TV and interacting with other viewers. This combination removes the need of physical co-location for social interaction. Thus, watching television online does not necessarily reduce one’s opportunities to build out their social networks.

Early research on the effects of Internet usage on social relationships by Nie and Hillygus (2002) and Shklovski, Kraut, and Rainie (2004) focused on a displacement hypothesis of Internet usage. This hypothesis states that time spent online competes with time spent with others face-to-face (Nie & Hillygus, 2002), for example by decreasing the likelihood that one visits a friend or family member (Shklovski, Kraut, & Rainie, 2004). These studies suggest that increases in Internet use lead to decreases in social capital because one has fewer in-person social encounters. This argument holds in the case of Internet users who engage in socially-isolating activities, such as playing single player games or watching videos alone. However, this argument may not hold for people who participate in online activities that involve communicating with others.

Prior research has shown that communicating with others online can lead to increases in social capital. In a study of Facebook users, Ellison, Steinfield, and Lampe (2007) found a positive association between students’ intensity of Facebook usage (defined as the frequency of visiting Facebook and emotional attachment to Facebook) and their social capital. In fact, this association was present both for bridging social capital – the extent to which the students felt integrated with their

campus community and their willingness to support their community – and bonding social capital – the maintenance of pre-existing close relationships.

McKenna, Green, and Gleason (2002) performed several studies of people who participated in a Usenet forum. They found evidence that people were able to form relationships in a completely online setting, without any face-to-face contact at all. In a longitudinal follow-up study, they also found that these relationships were stable over time. Ren, Kraut, and Kiesler (2007) discuss several main causes of how relationships are formed between people online. These causes include having social interactions with others and having personal knowledge of them. Frequent social interaction is associated with increases in liking (Cartwright & Zander, 1953), and is important for establishing bonds because it provides people with more opportunities to build social connections and create liking and trust. Social capital is built as a consequence of increases in communication and trust (Resnick, 2002). In addition, attachment to others increases when people have a sense of virtual co-presence or a subjective feeling of closeness with others in a virtual environment (Slater, Sadagic, & Schroeder, 2000). Self-disclosure, the act of revealing personal information about one's self, is another cause (and consequence) of interpersonal bonds (Collins & Miller, 1994). Therefore, to understand whether collaborative watching is an activity that can promote interpersonal bonding, and hence social capital building, this dissertation examines the ability of collaborative watching to increase subjective feelings of liking and closeness (Chapter 8), and whether those who chat while watching videos share personal details with each other (Chapters 8 & 9).

The argument for collaborative online video is that the videos provide an activity in which people can engage while they socialize with others. Indeed, watching television and conversing with others are activities enjoyed by many people around the world. However, it is unclear whether this activity of chatting with others while watching videos will have the same positive effect on social capital as has been seen in other online activities, such as playing massively multiplayer online games (Steinkuehler & Williams, 2006). Television is often an immersive experience (Lee & Lee, 1995), and people may not be interested in or capable of multitasking between watching it and socializing with others. Therefore, we must first understand the extent to which the immersiveness of the videos interacts with the ability of social features to provide a sociable experience. By sociable, I mean the extent to which a

viewer feels the presence of other viewers and enjoys interacting with them while watching.

This dissertation focuses specifically on the sociability of collaborative online viewing. Sociability is a prerequisite to social capital, as people cannot form relationships with each other if they do not feel each others presence in the online space. Since the prior work discussed has shown linkages between online communications and social capital, this dissertation argues that collaborative online video watching is a sociable activity, and by transitivity, can lead to gains in social capital with repeated, longitudinal participation. This argument is summarized below.

The sociability argument. Collaborative online video watching is a sociable experience that provides viewers with feelings of mutual connection to each other. Viewers enjoy chatting with others while watching videos together in an online setting. The videos act as a “social glue” that brings people together and provides them with enough common ground to bootstrap their conversations.

In this dissertation, I present evidence in support of the sociability argument.

1.3. HUMAN FACTORS AND HUMAN ATTENTION

Research in human factors and human attention demonstrates limitations in peoples’ ability to simultaneously process two or more sources of information. Multiple resource theory (Wickens, 2002) defines three stages in information processing: perception, cognition, and responding. Each of these stages are applicable to the visual (V) and auditory (A) modalities. Time-sharing between the modalities (AV) is generally easier than time-sharing within a modality (VV or AA; e.g., Wickens, Sandry, & Vidulich, 1983; Parkes & Coleman, 1990). Therefore, for example, we can we can look at a picture and listen to someone speak simultaneously without much interference. Trying to look at two pictures at once or simultaneously listen to two people speaking is difficult. For two visual channels (VV), if they are far enough apart, there is also a cost associated with the visual scanning required to move between the items. For two audio channels (AA), people are generally only able to attend to one channel at a time (Moray, 1969; Wickens & Hollands, 2000). A general model of auditory attention (Norman, 1968; Keele, 1973; Wickens & Hollands, 2000) proposes that input from the unattended auditory

channel remains in a short-term auditory store for about 3-6 seconds, and if a listener makes a conscious switch of attention, the contents of this store can be examined before it is lost. Despite this theory, dichotic listening tasks remain difficult and frustrating.

The activity of watching a video requires attending to the visual imagery with the eyes (visual perception) and processing that visual imagery into a meaningful representation of events (visual cognition). It also requires listening to audio of music, sound effects, and/or voice (auditory perception), and processing that audio into language (auditory cognition). In the case of collaborative watching, the chat feature overloads the attentional and computational resources in the brain, depending on the chat media used. For example, reading a text chat requires visual perception and cognition, which interferes with the visual perception and cognition required to watch the video. Voice chat requires auditory perception and cognition, which interferes with the auditory perception and cognition required to listen to the video's audio track. Thus, chatting and watching a video both compete for the exact same cognitive resources. Combining both activities seems doomed to failure. Interesting video would be missed when one is attending to chat. Interesting chat, or opportunities to respond to chat, would be missed when one is attending to the video. Therefore, collaborative watching may not result in an enjoyable experience because of the distraction present in multitasking between the video and the chat. This argument is summarized below.

The distraction argument. The human factors and human attention literature demonstrates that multitasking between two visual or two auditory channels is difficult because the two channels interfere with each other. This interference causes distraction. Distraction has two components: affectively, it causes a negative shift in one's mood because of the difficulty in maintaining attention to multiple sources of information; objectively, it causes one to miss information in one or both channels. Therefore, collaborative online video watching may not be able to engage viewers in an enjoyable and sociable experience because viewers will experience distraction from the combination of video and a chat feature, and this distraction will negatively impact their level of enjoyment.

In this dissertation, I present evidence that chatting while watching is distracting, but generally not to the point where it has an impact on viewers' enjoyment.

1.4. THESIS STATEMENT

The Oxford English Dictionary defines *collaborate* in the following manner:

col•lab•o•rate (intr. v.) \kə-ˈlɑ-bə-rāt\ : to work in conjunction with another or others, to co-operate

This dissertation is about the activity of *collaborative online video watching*, in which viewers work with each other to create a *social* experience while watching videos together online. They create this social experience by interacting with one another around the videos they watch. These interactions can occur among people who know each other very well or people who do not know each other at all.

In this dissertation, I focus on the case of watching and chatting in real-time. Although many online video sites provide social features that enable viewers to interact in an asynchronous fashion (e.g., commenting on or rating videos), synchronous communications are more intimate (Powazek, 2002), and seem more capable of providing a sociable experience.

The research questions addressed in this thesis revolve around sociability, distraction, scale, and learning in collaborative online video watching. In Part II, I demonstrate, through a series of laboratory and field experiments, that collaborative watching is sociable despite it being distracting. In Part III, I focus on the design and evaluation of user interface features that support sociable experiences in large-scale online video events like the one discussed at the beginning of this chapter. In Part IV, I demonstrate that collaborative watching has a productive quality to it, by allowing us to learn useful data about videos such as tags and ratings. This information can be mined directly from the chat logs produced by viewers who watch collaboratively.

My thesis statement is thus:

Collaborative online video watching – the activity of watching videos online while simultaneously chatting with one or more friends or strangers – is fun and sociable despite it being distracting, it scales to very large audiences, and it allows us to learn about videos from raw chat data.

1.5. CONTRIBUTIONS AND SIGNIFICANCE

This dissertation examines multitasking in an entertainment domain. Much work in human-computer interaction is focused on designing user interfaces that are easy to use, aid in the user's productivity, and do not frustrate the user. This dissertation demonstrates that, for user interfaces that combine video with chat, many users are willing to endure the distracting effects of the chat feature on the video in order to have a social experience with others. This dissertation also studies the effects that communication technologies have on viewers' enjoyment of online video, and finds that social interactions can improve enjoyment of poorer video content. This dissertation presents and evaluates designs for user interfaces that help users find and interact with each other during large-scale, real-time events. Finally, this dissertation demonstrates that useful information about videos can be inferred from the social interactions that take place among viewers in a large-scale audience.

This work will have a broad impact on the design of services and applications for online video. The results and design recommendations made in this dissertation will allow us to design better online video sites that provide viewers with more opportunities for social interaction.

Online video would not be possible without the high-bandwidth network links and content delivery mechanisms developed through research in computer networking. Although this dissertation is not focused on any part of the core networking technology that moves bits of data around the Internet, it does examine one of the most compelling applications for high-bandwidth networks to date: online video. This dissertation provides an additional value proposition for peer-to-peer video distribution systems. For content producers and distributors, value is derived from the cost savings in distributing video data as long as content consumers are willing to contribute their resources and share that cost. For users, the opportunities for social interaction in a peer-to-peer network with millions of nodes are overwhelming. These opportunities may provide users with additional incentives to participate in the network and share the costs of video distribution.

Prior research in human computation has sought to perform difficult or impossible computation tasks by engaging people in casual games. Sometimes these games are played alone or with a bot, and sometimes these games are played against other people. In each case, social interaction among players is actively prevented to avoid

collusion and prevent cheating. This dissertation demonstrates that human computation can be performed in an open social space in which people are free to interact with one another. The benefits to playing human computation games may be greater in a social context as well. Playing games is a solitary experience that may isolate one from others, as television has, whereas interacting with others promotes building and maintaining social capital.

The high-level contributions made by this dissertation include:

- Evidence that the distraction from multitasking does not always have negative consequences,
- Designs and evaluations of user interfaces for large-scale video broadcasts,
- Evidence that chat data collected from unmoderated social interactions can produce useful information about videos.

Specific contributions made by this dissertation include:

- A framework for understanding the different aspects of collaborative online video sites and how different design decisions affect the sociability of the experience (Chapter 3).
- Evidence that online video viewers want and already have social experiences around video media (Chapter 4).
- A methodology for examining collaborative online video watching in a controlled context, and scales for measuring experiential enjoyment, distraction, and sociability (Chapters 5-10 & Appendix B).
- Evidence that chatting while watching is fun and enjoyable for groups of friends and groups of strangers (Chapters 6, 8, & 9).
- Evidence that chatting while watching is distracting (Chapters 6-9) and that short break periods between videos alleviates distraction (Chapters 8 & 12).
- Evidence that chatting while watching poorer content improves enjoyment of that content (Chapter 8).
- Evidence that voice chat is no more distracting than text chat, and that viewers express a preference for voice chat when exposed to it (Chapter 9).

- Evidence that relevant information about videos can be extracted from chat data and that social interactions can be used to perform human computation tasks (Chapters 13-16).
- Design and evaluation of a novel social proxy for representing large, virtual audiences (Chapter 11).
- Evaluation of tag clouds and scrolling lists of messages as visual summaries that provide awareness of a large audiences' chat activities (Chapter 11).
- Motivation for the design of new P2P video streaming protocols that leverage social networks to relax strong synchronization requirements in order to reduce loss rates and improve the quality of the video (Chapter 11).
- Design of a collaborative online video watching site that supports watching videos with friends on Facebook (Chapter 12).

1.6. OVERVIEW

This dissertation consists of five parts. Part I establishes the need for research in collaborative online video watching and discusses related work in television, computer-mediated communication, online communities, human factors and attention, social television, and online video. Chapter 3 presents a framework for the design of collaborative online video sites and examines how different design options affect the scope of social interactions on the site. Part I concludes with a survey of YouTube users demonstrating that online video is not only popular, but often experienced in a social context.

Part II discusses four empirical studies I have run that examine the core experience of collaborative online video watching – chatting while watching videos with others – from the perspective of small groups watching together in a controlled laboratory environment. These studies establish that chatting while watching is fun, enjoyable, and sociable. Several of these studies explore the issue of distraction that arises when combining two activities – chatting and watching – that compete for a viewer's attention (Chapters 7-9). I discuss a strategy for reducing distraction in Chapter 8 using short break periods in between videos. In Chapter 9, I discuss a media comparison between textual and auditory chat, and show that despite our intuition that voice chat would increase distraction, it did not. Part II concludes with a general discussion of the findings across all of the studies in Chapter 10.

Online video events, such as the inauguration of President Obama, are becoming increasingly popular and attracting audiences upwards of millions of viewers. Part III considers the design of user interfaces that help people find enjoyable social experiences whilst watching a video broadcast in a large audience. Using features like visual chat summaries and social proxies (Chapter 11), I present a design and implementation of a collaborative online video watching application in Chapter 12. This application combines the social networking features of Facebook with the online video library of YouTube to enable both friends and strangers to chat with each other while watching online videos. I present results from a study of the initial version of this application, as well as mockups and screenshots of its redesign.

In Part IV, I consider how collaborative online video watching can be used to improve the quality of online video sites by mining chat data for useful information about videos. This information includes a set of tags that can be used to label a video (Chapter 14), hints about people's enjoyment of a video that can be used to improve the accuracy of video ratings (Chapter 15), and moment-by-moment profiles of videos that show which parts viewers most enjoyed (Chapter 16). Part IV establishes collaborative online video watching as a human computation task in which viewers provide labels and ratings for videos indirectly, as by-products of their social interactions. Such labels and ratings either cannot be inferred by traditional computational tasks, or can only be inferred with questionable accuracy.

I conclude this dissertation in Part V by discussing limitations and future work (Chapter 17) and the general conclusion that unlike television, online video need not be an isolating experience (Chapter 18).

As an aid to the reader, Table 1-2 summarizes the nine studies detailed in this dissertation. Figure 1-2 provides a visual diagram of these studies that depicts the flow of ideas over time between studies, analyses, and interface designs.

Table 1-2. The studies discussed in this dissertation.

Laboratory	Real-world	Evaluation
<i>Live and simulated studies</i>	<i>Studies of real-world systems</i>	<i>Rating and think aloud studies</i>
Chat Distraction (Ch. 7)	MovieLens (Ch. 6)	Best Part Labeling (Ch. 16)
Cartoon (Ch. 8)	YouTube survey (Ch. 4)	Tag Evaluation (Ch. 14)
Text vs. Audio (Ch. 9)	Social Video (Ch. 12)	Large Audience (Ch. 11)

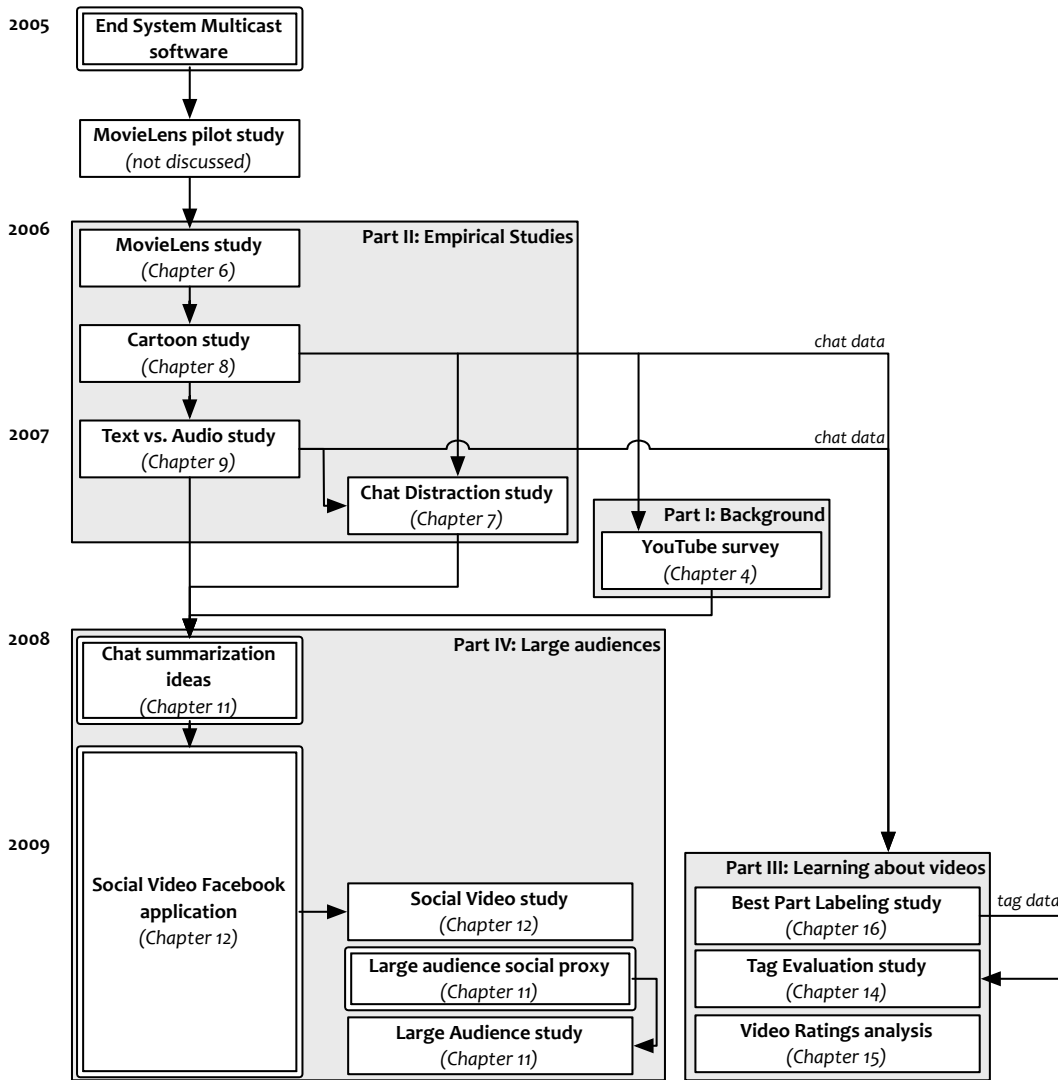


Figure 1-2. Timeline of the studies and design projects presented in this dissertation and their corresponding chapters. Arrows represent the approximate flow of ideas and influence between studies as well as data when noted. The height of the boxes is generally not reflective of the amount of time spent on each project. Boxes with a double border represent projects focused specifically on design and implementation.

Part I: Background

Is chatting while watching videos online usable, or is it too distracting to multitask between these two activities? Research in human factors and human attention suggests chatting while watching will be distracting. In Part

I, I discuss research in computer-mediated communication and social television that encourages the use of a chat feature to promote sociability. I also present a framework for the design of collaborative online video sites that highlights how different design decisions can affect sociability. Finally, I motivate the need for research in social viewing by presenting data from a survey of the social and sharing behaviors of YouTube users.

2.

COMBINING CHAT WITH VIDEO: RELATED WORK

Is chatting while watching video usable? Will viewers be able to multitask between the two activities? Although the human factors and human attention literature discussed in Chapter 1 suggests that chatting while watching will be distracting, prior research on television-watching shows that people already multitask while watching TV. Research in computer-mediated communication has examined chatting online and the ability of chat features, such as text chat and voice chat, to support back-and-forth conversations and convey a sense of social presence. Finally, chat features have been added to many other online applications as well, such as distance learning and remote presentations. These applications are similar to collaborative online video watching as they require users to multitask between the activity and the chat feature. Thus, prior research suggests that chatting while watching may indeed be usable. This chapter concludes by summarizing current research in social and interactive television, a nascent field focused on designing collaborative online viewing experiences for both the computer and the couch.

2.1. TELEVISION

Research on television-watching and television audiences has been conducted to understand why people watch television, how they watch it, and how watching it with others affects their experience. Clancey (1994) addresses questions of what constitutes television “watching,” why people watch TV, and how they watch. She defined “watching” behaviorally, by being in a room with a TV set turned on. She

found that people were more likely to watch alone during the morning and afternoon, and with other people during the evening and prime-time hours. She also found that people do multitask while watching by performing activities such as eating, reading, talking, doing homework, and resting.

Lee and Lee (1995) also examined television-watching behaviors with a particular focus on how peoples' uses of television would impact the development of interactive television services. They view social television watching as an impediment to providing individualized television content and services, although they found that roughly two-thirds of prime-time viewing was done in the company of others. They also found that people watched TV to improve their mood and help them relax, to stay up-to-date on current affairs, to learn about how to cope with different life situations, to escape from reality, and to have a experience they can share with others by watching together and/or talking about television programs. Interestingly, Lee and Lee (1995) predicted that the convergence between television and PC was unlikely, and instead both devices would compete with each other for the "prime-time attention of America" (Lee & Lee, 1995, p. 16). This prediction was borne out in the results of a study by Kraut et al. (2004), which found that increased use of the Internet was associated with declines in television watching.

Watching content with others is associated with increases in enjoyment. Raghunathan and Corfman (2006) performed a study in which participants watched ads with a confederate. The confederate either expressed similar, neutral, or different opinions about the ads as the participant. They found that participants enjoyed the ads more when the confederate expressed similar opinions, and they concluded that promoting interactions among participants with similar reactions to an experience increases their enjoyment of the experience and encourages them to repeat it. Consistent with these results, a study by Ramanathan and McGill (2007) found that viewers who watched a video clip in the presence of another person engaged in a non-conscious mimicry pattern in their moment-by-moment ratings of that clip. They concluded that watching with another person affects one's moment-by-moment reactions to be more in line with those of the other person.

Vosgerau, Wertenbroch, and Carmon (2006) studied peoples' attitudes toward watching indeterminate content, where the outcome is unknown ahead of time, such as a football match. They found that watching live was associated with greater levels of excitement and preference. They also found that people were more likely to watch

with others when watching live, although their anticipated enjoyment of watching live material did not change if they watched it alone or with others.

Research in television-watching shows that people do enjoy watching as a social experience, and suggests that collaborative online video watching may be sociable as well.

2.2. COMPUTER-MEDIATED COMMUNICATION AND MEDIA RICHNESS

Computer-mediated communication is a field devoted to studying how humans communicate using technology. Often, these communications occur in a remote context in which the people communicating with each other are not physically co-located. Challenges that computer-mediated communications face include the ability to convey communicative cues, the ability to promote trust and cooperation, and the degree to which communications partners can establish common ground. Without these abilities, it is difficult for communications partners to communicate efficiently and effectively online, impeding sociability. Prior work shows, however, that both text and voice chats can be effectively used for collaboration.

When people communicate with each other, they do so through some channel. Media richness is the notion that communications media differ in their ability to facilitate the creation of shared meaning among communications partners (Daft & Lengel, 1984; Daft & Lengel, 1986; Dennis & Kinney, 1998; Walther & Parks, 2002). They assert that four factors influence the richness of a medium: the ability to transmit multiple cues (e.g., vocal inflection or gestures), the immediacy of feedback (e.g., synchronous vs. asynchronous), the variety of language supported (e.g., formal vs. conversational), and the degree to which messages can be personalized (e.g., tailoring messages to a specific individual vs. a large audience). In general, richer media provide more support for these factors than leaner media.

The initial studies of media richness by Daft and Lengel (1983) considered media richness in a workplace context, and discussed communications media commonly used in offices of the time: face-to-face, telephone, written communication, and computer output. More recently, media richness has been applied to computer-mediated channels to describe their ability to convey non-verbal cues (Dennis & Kinney, 1998). These channels generally take the form of text, voice, and/or video.

Computer-mediated communication researchers have studied and compared these chat media to understand the contexts in which their use is appropriate or helpful, as well as their limitations in supporting online groups. In the workplace, instant messaging is often used to check the availability of others (e.g., Nardi, Whittaker, & Bradner, 2000; Isaacs et al., 2002), to conduct complex work-related conversation (e.g., Isaacs et al., 2002), to initiate interactions in other, richer media (Connel et al., 2001), and even socialize (Weisz, Erickson, & Kellogg, 2006). Group text chats have also been used to support work groups, and they capitalize on the fact that text chat supports both synchronous and asynchronous interactions (Handel & Herbsleb, 2002; Erickson et al., 1999). However, text chats are not without problems. Smith, Cadiz, and Burkhalter (2000) identify five problems with text chat, outlined in Table 2-1. In their work, they attempted to overcome these issues by designing a new user interface for text chat that organized messages hierarchically, showing the explicit message/reply structure. However, participants in their user study found it difficult to follow the conversation because it was hard to find new messages as they were added.

Table 2-1. Five core problems with text chat, adopted from Smith, Cadiz, and Burkhalter (2000).

Problem	Description	Resolution(s)
Lack of links between people and what they say	Chat interfaces make it difficult to differentiate speakers, e.g., by only associating messages with the name of their speaker	Addressable by making individual participants visually distinguishable from each other (e.g., Viégas & Donath, 1999; Vronay, Smith, & Drucker, 1999)
No visibility of listening-in-process	Chat participants do not receive moment-by-moment indications of whether others are listening; the lack of a listening cue reduces social presence	Social proxies indicate active participation, such as composing a new message, but not necessarily listening (Erickson et al., 1999)
Lack of visibility of turns-in-progress	Messages are only transmitted when the return key is pressed; chat is not truly synchronous, making turn-taking harder	Fully synchronous chat shows letters as they are typed; a typing indicator is commonly used in instant messaging applications to aid turn-taking

Problem	Description	Resolution(s)
Lack of control over turn positioning	Chat messages are organized temporally, not topically; fast-paced chats often have messages that respond to a message that occurred more than one turn ago	Slowing the pace of chat (e.g., Erickson et al., 1999); displaying chat messages in a thread structure (Smith, Cadiz, & Burkhalter, 2000)
Lack of useful recordings and social context	Chat groups do not accrete a social history; transcripts are difficult to comprehend	Threaded chat produces a browsable history of chat (Smith, Cadiz, & Burkhalter, 2000)

In addition to the problems identified by Smith, Cadiz, and Burkhalter (2000), O'Neill and Martin (2003) performed an analysis of text chat conversations and found that schisms in the conversation commonly occurred. Schisms are points at which the topic of a conversation branches into two separate subtopics ("conversational threads"), which may be continued independently or merge back together. Schisms were also identified by Sacks, Schegloff, and Jefferson (1974) in face-to-face conversations, and tended to occur more frequently in groups larger than three or four members.

Despite these "core" problems with text chat, there are benefits as well. Early work by Kiesler, Siegel, and McGuire (1984) found that text chat usually fosters equal participation in groups. Groups using text chat can find common ground when working remotely (Birnholz et al., 2005). Text chat helps workers get instant access to their questions (Weisz, Erickson, & Kellogg, 2006), and supports the formation of personal relationships online (Parks & Roberts, 1998; McKenna, Green, & Gleason, 2002).

Voice chat has often been compared with text chat because it is a richer medium. Voice chat is able to convey vocal inflection and tone, allowing message senders to add additional cues to their messages. In general, voice chat has been shown to foster greater levels of trust and cooperation than text chat in social dilemma tasks (Bos et al., 2002; Jensen et al., 2000), and it fosters greater levels of trust and liking than text chat in multiplayer gaming (Williams, Caplan, & Xiong, 2007). However, Löber, Schwabe, and Grimm (2007) observed productivity losses for larger groups (seven members) using voice chat, and productivity gains for larger groups using text chat. Scholl, McCarthy, and Harr (2006) argue for the inclusion of text chat with a voice and video channel. In this combination, text chat enables asynchronous

communication and can compensate for difficulties in audio. The voice and video features provide a richer indication of presence and emotional feedback. Geerts (2006) notes that age may make a difference in media preference; in his study of social TV users, voice chat was preferred overall, but text chat was preferred by younger users and those with more experience chatting on computers.

In conclusion, chatting through computer-mediated channels is generally more difficult and less efficient than chatting face-to-face. However, despite the problems with computer-mediated chats, users are still able to collaborate, establish common ground, and build trust with one another.

2.3. APPLICATIONS OF CHAT WITH VIDEO

Simultaneous chat with a source of video has been studied in several contexts prior to online video watching. This section summarizes prior work in combining chat with video to create distance learning, remote presentation, and video conferencing applications. It concludes by summarizing work in the area of social television, a nascent field devoted to designing for and studying social behaviors around television and online video.

2.3.1. DISTANCE LEARNING

Distance learning applications often combine a video lecture with a computer-mediated chat channel for students. Leonard, Riley, and Staman (2003) summarize the technologies used to create virtual classroom settings online, and describe the capabilities of different interactivity features such as shared whiteboards, instant messaging, and email. They conclude that (at the time), although the technologies exist to support the creation of virtual classrooms, a large challenge is to motivate teachers to creatively exploit those tools to enhance the classroom experience.

Flatland creates an environment for distance learning by combining live, streaming video with slides and interactivity features such as text chat, a question queue, and quizzes (White et al., 2000). In an evaluation with students, White et al. (2000) found that students multitasked during the slower parts of lectures, but recognized when to pay attention using verbal cues from the instructors. Instructors initially felt that in-person lectures were superior to the virtual classroom. Over time, their comfort with the system increased, and they used more of the interactivity features

such as the Q&A window and dynamic text slides to type in code examples. One issue that hindered usability was the video latency between the instructor and the students. In some cases, this latency caused instructors to miss questions from students because they did not wait long enough when asking if students had questions. In general, instructors requested more awareness of their remote students through audio or video channels.

The Interactive Shared Educational Environment (Mu, Marchionini, & Pattee, 2003) seems to be a direct response to this challenge. The ISEE combines a video player with an integrated text chat channel and a shared web browser component to create a distance learning environment. In a user study, they found that participants were comfortable using this environment, although text comments that did not synchronize to the video were somewhat distracting.

Another distance learning system that combines chat with video is modeled after Tutored Video Instruction (TVI; Gibbons, Kincheloe, & Down, 1977). In TVI, students watch a videotape together on a particular subject. At certain points in the video, the instructor pauses the video and leads a short discussion of the material seen so far. In their study, Gibbons, Kincheloe, and Down (1977) found that students performed better on an exam using the TVI method compared to students who watched the lectures in the classroom, and to students who watched the lectures individually. This finding was replicated by Stone (1990).

In Distributed TVI (DTV; Cadiz et al., 2000), students perform the TVI process in an online setting. In a DTVI study conducted by Smith, Sipusic, and Pannoni (1999), students discussed lecture material with each other using high-quality, low-latency audio and video links. As with the previous TVI studies, the discussion fostered by the TVI method had a positive impact on student grades.

2.3.2. PRESENTATIONS AND LECTURES

One early system that supported distributed presentations was Forum (Isaacs, Morris, & Rodriguez, 1994). In Forum, presenters communicated to a remote audience using voice and video. It contained several features that enabled speaker-to-audience, audience-to-speaker, and audience-to-audience interactions. These included a question-asking queue (using a “raise your hand” metaphor), live polling of the audience, and sharing and annotating slides. Speakers reported that voice chat

was important because it gave them a better understanding of the audience and their interest level. Audience members enjoyed seeing the video of the speaker because it felt more intimate and appeared as if the speaker was talking directly to them. In a comparison of remote and face-to-face presentations, Isaacs, et al. (1995) found that attendees enjoyed attending talks remotely with Forum because it enabled them to multitask and archive the presentation materials (e.g., slides), and it was more convenient attending remotely than in person.

To further tie together local and remote audiences, Jancke, Grudin, and Gupta (2000) designed the TELEP system to increase the awareness of remote audiences during lectures and talks. Their system placed a display of the faces of remote attendees – either as a static picture or a live video feed – in the lecture hall. In addition, remote attendees could ask questions of the speaker using a question feature, and could chat with each other using a text chat feature. Both speakers and local audience members found the TELEP system interesting, and experienced an increased awareness of remote audience members. Remote audience members tended not to stay in the lecture as long as local audience members, and reported focusing on the talks about half of the time; the rest of the time was taken by doing other work, thinking, or daydreaming.

Baecker (2003) summarizes the design choices present in creating webcasting systems and services. The requirements for a webcasting system fall into five categories: participants, media, interactivity, archives, and system. In general, he recommends that webcasting systems should support both local and remote audiences, prioritize audio delivery over video delivery, provide interactivity features for remote audience members, and archive talks so that users can view them (non-linearly) after the talk has ended. Baecker (2003) uses these design recommendations to design ePresence, a system for running and archiving distributed presentations and lectures. In his evaluation, Baecker reports that users were generally satisfied with ePresence. A separate study of ePresence (Baecker et al., 2006) found that users felt isolated and least engaged when unable to chat with each other, and most engaged when using a voice chat feature.

2.3.3. VIDEO CONFERENCING

Video conferencing has been used for many applications, including tele-psychiatry (e.g., Pesämaa et al., 2007), language teaching (e.g., Hampel & Baber, 2003), and

team collaboration (e.g., Kauff & Schreer, 2002; Fish, Kraut, & Root, 1992; Fish et al., 1993). Focusing on the latter case, CRUISER was a video conferencing system designed to promote informal communications in the workplace (Fish, Kraut, & Root, 1992; Fish et al., 1993). In their evaluations, Fish et al. found that the open video channel provided users with background awareness of each others' status and minimized the cost for starting conversations. However, they also found that conversations over the video channel were shorter than face-to-face conversations, and users felt that CRUISER invaded their privacy when other users initiated video conferences with them. A feature designed to promote informal communication by randomly initiating voice conference calls (named "autocruise") was reported as being overly intrusive. Therefore, Fish et al. concluded that the CRUISER system was generally unable to support informal communications. It was difficult for users to emulate the face-to-face conversational protocol for initiating informal communication, and the system-supported "autocruise" mechanism was too intrusive to be useful.

2.3.4. SOCIAL TELEVISION

Research in social television is focused on creating systems that integrate social interaction during the television-watching experience. An early example of a social TV system is 2BeOn (Abreu, Almeida, & Branco, 2001; Abreu & Almeida, 2009), the design of which called for interactivity features that enabled viewers to communicate through instant messaging, voice chat, and video conferencing. AmigoTV is another early social TV system, which combines voice chat with an overlay display on the TV (Coppens, Trappeniers, & Godon, 2004). AmigoTV shows avatars of other viewers on the screen, and it shows each viewer what their friends are watching. An evaluation of AmigoTV by Geerts (2006) found that participants enjoyed using the voice chat feature, although they felt it was distracting to simultaneously watch and chat.

Media Center Buddies combines instant messaging and television using a Media Center PC (Regan & Todd, 2004). In their user study, they found a significant increase in enjoyment when viewers used the IM feature with their friends while watching. Another instant messaging application is Reality IM (Chuah, 2003), which uses a different approach than Media Center Buddies. In Reality IM, a bot is used to provide information about the current television program being watched, such as

statistics for a golf player. In addition, it can be used to enable commerce by allowing viewers to purchase the products they see in advertisements and in the program directly from the IM session.

Other social television systems use different approaches to promote sociability. Telebuddies encourages interaction through quiz games and trivia contents during television broadcasts (Luyten et al., 2006). STV1 allows friends and family to chat with each other while they watch a program together using open microphones in the living room (Harboe et al., 2008a). Its successor, STV2, adds an ambient display to inform users of when their (remote) friends are currently watching. (Harboe et al., 2008b; Harboe et al., 2009). The CollaboraTV interface represents friends with avatars and shows them watching together in a virtual audience (Harrison & Amento, 2007; Nathan et al., 2008; Amento et al., 2009). ConnecTV (Boertjes et al., 2009) allows friends to send recommendations to each other for programs to watch.

One difficulty designers face in creating social TV systems is in determining who is watching the same program at the same time. The previous systems discussed generally use a viewer's existing instant messaging contacts to show which friends or family members are online and watching. Research by Fink, Covell, and Baluja (2006 & 2008) demonstrates how to use audio fingerprinting to determine the show a viewer is watching. In scale, this method can be used to create ad-hoc communities of all viewers who are watching the same program simultaneously. This idea was implemented in the Cha.TV system (Fink, Covell, & Baluja, 2006). Thus, social television need not be restricted to well-defined groups of friends or family. Audiences of strangers can also be composed on-the-fly, and information about their presence can be displayed to make watching television feel more social.

2.4. SUMMARY AND CONCLUSIONS

- Watching television is often done socially, and viewers often multitask while watching by talking, doing homework, or even resting. Watching in the company of others is associated with increased enjoyment, and watching live broadcasts is associated with greater levels of excitement. Research in television-watching supports the usability of collaborative online video watching, as viewers are already used to multitasking and socializing while watching.

- Computer-mediated communication research provides a framework for understanding how different communications channels affect the ability for remote viewers to effectively communicate with each other. Although researchers have identified difficulties users have in communicating with others online, research has shown that common ground, trust, and relationships can all be established between remote partners.
- The combination of video and chat has been studied in other venues, including distance learning, remote presentations, and video conferencing. Chatting while watching online lectures is associated with learning gains, although lecturers and presenters generally found it difficult to maintain awareness of their remote audience. Open video channels can provide awareness of a remote user's status, albeit in an intrusive manner. These examples highlight the fact that combining chat with video is not always a trivial matter.
- Researchers in social and interactive television are working to transform an isolating television-watching experience into a socially-engaging one by incorporating features that enable remote viewers to see when they are present, chat with each other, and share and recommend shows to watch. In their evaluations of social television systems, they have found that viewers generally enjoy chatting while watching, and that chatting while watching is distracting. The studies presented in Part II of this dissertation perform a more rigorous analysis of sociability and distraction.

3.

A FRAMEWORK FOR COLLABORATIVE ONLINE VIDEO²

There are many different ways to experience video online. Some online video sites and Internet video applications offer videos of a particular genre or theme; others contain videos from many genres. Some allow only short video clips; others offer full-length movies. Some encourage viewers to interact with each other before, during, or after the activity of watching videos; others do not. These different options constitute design decisions that can affect the entertainment and social value of an online video site.

In this chapter, I present a framework for designing collaborative online video experiences. This framework highlights different aspects of the collaborative viewing experience, including what viewers watch, where they watch it, and with whom they interact while watching. Where appropriate, it also details how different design choices can impact sociability among viewers, and ultimately, the ability for the site or service to build a community. Figure 3-1 shows a schematic diagram of the five aspects of a collaborative online video site discussed in this chapter.

Many online video sites are used as examples in this chapter to highlight different design options. For reference, these sites are described in more detail in Table A-1 in Appendix A.

² Portions of this chapter have previously appeared in (Weisz, 2009).

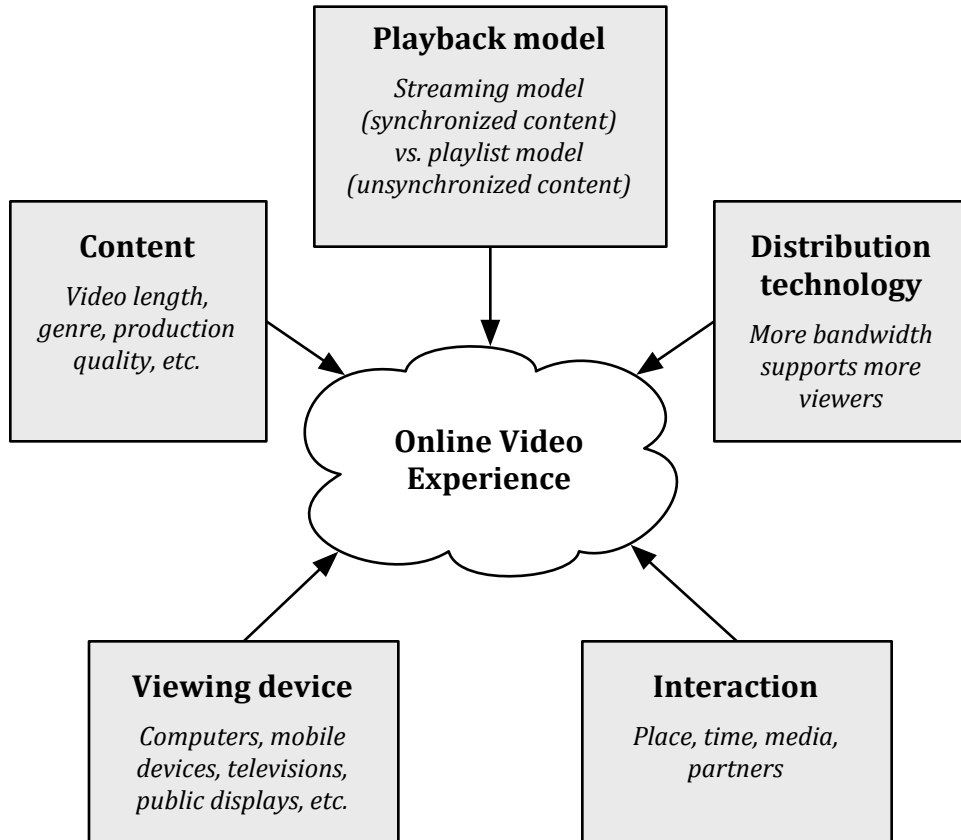


Figure 3-1. Five aspects of the collaborative online video experience.

3.1. CONTENT

Many different types of videos are offered online. Video genres include cartoons, music videos, political speeches, “reality TV,” and many others. Videos also vary in running time, and some sites have placed explicit limits on the length of content. For example, YouTube initially limited uploaded videos to about 10 minutes in length; in 2008 they rescinded this limit (Graham, 2008).

The purpose of the site and its video content also varies. Some sites host videos created purely for entertainment or humor, such as videos typically hosted on CollegeHumor.com or Break.com. Other sites create or encourage videos for informational, educational, or political purposes, such as the 2008 CNN/YouTube debate partnership, talks on TED.com, and the video lectures of MIT’s OpenCourseWare.

Video sites also vary in the production quality of the videos they present. Some sites focus on professionally produced videos having pre-written and rehearsed scripts, or having professional equipment and lighting used during production. Examples of these sites include online TV shows like *The Scene*, *PurePwnage*, and *Red vs. Blue*, as well as the *Rocketboom* news podcast. Other sites present videos taken on the spur of the moment, which are often recorded with low-fidelity equipment such as a webcam or cell phone.

These differences in video content are important to understand because they serve to define the type of community an online video site fosters. Video content affects the types of viewers the site attracts as well as the interactions that take place among those viewers. For example, a site that presents political and newsworthy content tends to attract an audience interested in politics, and will promote political conversation. A site that specializes in short, amateurish, and humorous video clips may promote a myriad of conversational topics, including nonsensical or humorous remarks and banter. In this dissertation, I utilize brief entertainment content similar to that found on YouTube, because entertainment videos comprise the most popular segment of the online video space, and because social experiences often take place in an entertainment context.

3.2. DISTRIBUTION TECHNOLOGY

Online video sites employ two primary models of video delivery. The most common is the download model, whereby viewers download videos and/or watch them as they download. This is the model currently used by sites like YouTube and Yahoo Video. This model gives viewers the flexibility of watching a video at some future time and of watching on a portable device such as a video-enabled iPod.

The streaming model is familiar from classical (DVR-less) television. In this model, a video is streamed live as the viewer watches it; tune in during the middle of the stream and you've missed the beginning. Sites such as Justin.TV and UStream.TV currently use this model.

Each of these models has implications for the entertainment and social value of an online video site. The download model allows for extended, asynchronous conversations. Viewers can post comments and get responses on any video they have watched. They can also chat in real time about the videos. The download model

also gives viewers greater flexibility in selecting content to watch. Since viewers are not constrained by having to watch the same video content as everyone else, they are free to explore a wider range of content, picking and choosing what to watch at their leisure. However, this model has implications for sociability. Since viewers do not necessarily watch the same content at the same time, synchronous chat may be less effective at providing a social experience for viewers, because they have less common ground than their streaming-video peers. For example, when watching a soccer match, a viewer who sees his team score a goal might want to cheer and celebrate, but those cheers may fall on confused ears if no one else is watching the same soccer match at the same time.

Compared to the download model, the streaming model is constrained by its synchrony. All viewers see the video stream near-simultaneously in time, although network delays may cause small asynchronies in playback. The most natural choice for integrating social interaction into this model is synchronous chat. This feature allows, for example, viewers of a soccer match to share their reactions and excitement as a player scores a goal the moment it happens. Sharing these moments is one important reason why people enjoy watching sporting events live, because events are revealed to the viewers together and they can share their reactions with each other (Vosgerau, Wertenbroch, & Carmon, 2006). An asynchronous chat medium would fail to capture such an experience.

Another aspect of video delivery relates to the technology used to deliver video data. The choice of technology has implications for the size of the viewing audience. Three of the common methods used to deliver video data are shown in Table 3-1. Each of these methods can be used to implement either a download or a streaming model.

The client-server model is the simplest and least expensive to implement. It works by having a server (or group of servers) deliver video data to each viewer. In this model, the size of the audience is limited by the amount of available bandwidth on each server. Because of this, the client-server model only supports a few hundred to a few thousand simultaneous users.

Table 3-1. Methods for online video delivery.

Distribution method	Description	Tradeoffs	Implications
Client-server	One or more servers send video data	Inexpensive; easy to implement; does not scale well	Smaller audiences; few opportunities to meet new people
Content distribution network (CDN)	Multiple servers used to send data; delivery optimized by geography	Expensive; difficult to implement; scales well	Large audiences; many opportunities to meet new people
Peer-to-peer (P2P)	Viewers distribute video data to each other	Inexpensive; difficult to implement; scales well	Large audiences; many opportunities to meet new people

Content distribution networks (CDNs) can be used to reach the next order of magnitude of users. CDNs are essentially multiple instances of the client-server model, but with one key optimization. Videos (or video streams) are replicated across many different servers located in different geographic regions, and video data are delivered to viewers from a proximate server. This allows the CDN model to scale to hundreds of thousands or millions of viewers. However, the costs to create and maintain such an infrastructure are high, and thus so is the cost of delivering video to a large audience. These costs are typically prohibitive to an individual user seeking to reach a large audience.

An alternative to using a commercial service is to use a free video hosting service such as YouTube or Yahoo Video. These sites allow members to upload videos for free and make them available for watching by millions of viewers. Underneath the hood, these services are also CDNs, in that video files are replicated across many servers and delivery is optimized by geography. However, this approach still requires a significant investment in bandwidth (on YouTube or Yahoo's part). Moreover, as of this writing, they do not currently support streaming videos to a large audience. Thus, while this solution does allow millions of users to download a video, it does not yet allow them to watch simultaneously. Further, these services are experiencing a pressing need to generate revenue in order to remain financially solvent. Their hope is to generate revenue by serving advertisements alongside their videos, but advertisements may not be acceptable to some users.

The peer-to-peer (P2P) model overcomes the limitations of the client-server and CDN models by distributing the work of delivering video data across all viewers.

Peer-to-peer models support both downloading and live streaming for video. Many different P2P protocols and systems exist, including End System Multicast, Sopcast, TVU, and PPLive. These technologies are beginning to enter into mainstream use (Grigonis, 2008) and they have the potential to revolutionize the distribution of online video by allowing individual users to broadcast video to audiences of thousands or millions. As P2P technologies become more popular, the potential for online video events with millions of users becomes more realistic. Thus, it is important to begin exploring the implications of an audience of this size and the challenges it presents to interaction among viewers. Part III of this dissertation addresses the challenges of interacting with other viewers in a large audience, and presents concrete designs for user interface features that help viewers understand who is in the audience and what they are talking about.

3.3. VIEWING DEVICE

The phrase “online video” has strong connotations that the video is, in fact, being viewed on a computer connected to the Internet. This connotation is somewhat misleading, as there are many different options for viewing “online video.” Although this dissertation focuses on watching videos on a computer, it is important to understand that online video is enabling new kinds of viewing experiences on the television, on computers, and on mobile devices.

Research in interactive and social television has focused on creating systems that integrate social interaction during the television-watching experience. AmigoTV combines voice chat with an overlay display on the TV showing avatars of other viewers, and allows users to see what their friends are watching (Coppens, Trappeniers, & Godon, 2004). Media Center Buddies allows viewers to send instant messages and chat with one another while watching television (Regan & Todd, 2004). Telebuddies encourages interaction by allowing content producers to add quiz games and trivia contests to their broadcast (Luyten et al., 2006). Reality IM combines social interaction with additional contextual information about a TV program, such as advertisements and sports statistics (Chuah, 2003). Social TV 2 uses an ambient display to inform users of when their friends are watching television, and it allows them to chat with each other when they are watching at the same time (Harboe et al., 2008b).

This idea of interacting around television content is spreading to online communities as well. Fan-based and officially-sponsored communities exist for popular TV shows and movies. Some of the major broadcasting companies in the US, including NBC, ABC, and CBS, have dedicated discussion boards online for their television programs. Fan-run sites like Cinema Blend and First Showing have discussion boards and blog commentary for the latest movies. Further, some communities have created a new viewing experience by offering live chat as a television show is aired, augmenting the television with a computer to transform a solitary viewing experience into a socially engaging one. One example of this is the Lostpedia community, which hosts live chats for fans of the show *Lost* as it airs.

Mobile devices are also becoming a popular way for watching online video. Cell phone providers offer streaming video services to video-capable cell phones using high-speed cellular networks. The iPhone and iPod Touch allow people to watch videos directly from YouTube on their devices wherever they are, through either wireless or cellular networks. Thus, these devices allow people to watch video on the go, whenever they want (time-shifting) and wherever they are (space-shifting). The proliferation of video-capable mobile devices will expand the number of design options for creating entertaining and social video experiences.

In this dissertation, I focus on the case of watching videos and chatting with others on a computer. This is a common case for online video, and less research has been done on the experience of watching and chatting on a computer as has been done on chatting with others while watching television.

3.4. INTERACTION

Four facets of interaction around online video include place, time, media, and partners. Place and time are concerned with where viewers are located as they watch and whether viewers watch at the same time (synchronously) or at different times (asynchronously) from each other. Media concerns the ways in which viewers interact with each other, such as using textual, auditory, or visual media, or some combination. Partners concerns the relationships between viewers, such as whether they are friends, acquaintances, strangers, or even anonymous or imagined.

Place and time. One way to understand the interaction space of online video is through the time-space taxonomy heretofore applied to groupware systems (Ellis,

Gibbs, & Rein, 1991). Table 3-2 describes four types of interactions among viewers that online video is capable of supporting, and gives an example of each.

Table 3-2. Time-space taxonomy applied to online video interactions.

	Same time	Different times
Same place	Face-to-face interaction; e.g., a group watches tennis on a mobile device or computer together	Asynchronous interaction; e.g., recording a video for someone else or leaving comments on a public video display
Different places	Synchronous interaction; e.g., dispersed friends watch a tennis match together and chat while they watch	Asynchronous interaction; e.g., dispersed friends watch a YouTube video and post comments to each other

In this time-space framework, viewers watching at the same time can synchronously chat with each other as they watch, whether or not they are in the same place. Those watching at different times can communicate only asynchronously.

Sites that use a download model for video delivery typically have a space where members can post messages or comments in an asynchronous fashion. However, it is also possible to add synchronous communication to a download model. For example, in YouTube Streams³, each “stream” is a chat room with its own video library. Viewers can build a playlist of videos they want to watch from this library. They can also chat with each other in real time while watching videos from their own playlist. This model blends synchronous interaction with unsynchronized content.

In this dissertation, I focus only on the case of remote, synchronous interactions. Synchronous interactions are more intimate than asynchronous ones (Powazek, 2002), and thus seem to be more capable of providing a sociable experience for collaborative online video watching.

Media. Online video sites can also differ in the type of media they use for interaction. Although many sites implement text-based interactions, both for live chat and posting messages or comments, some sites have been experimenting with alternative media. For example, Gaia Online allows its members to watch videos in a 2D graphical chat room. Viewers are represented as graphical avatars and chat is

³ YouTube Streams. http://www.youtube.com/streams_main

displayed using chat bubbles. Second Life, a 3D virtual environment, allows its members to embed videos in virtual televisions and movie screens. Chat in Second Life is also text-based, and uses both chat bubbles as well as a text box to display chat. Other chat media are also possible, such as audio or video chat while watching. These examples illustrate the wide variety of possibilities in designing interactive experiences for viewers.

In this dissertation, I consider the usability of both textual and auditory chat channels. From a media richness perspective, voice chats allow people to convey a broader variety of cues through tone and inflection. From a usability perspective, voice chats can make a chat feature more accessible to people who are uncomfortable with technology or unable to use computer keyboards. The comparison between text and voice chat is presented in Chapter 9.

Partners. Chatting while watching a video can take place among people with a broad range of pre-existing social relationships. For example, viewers can be best friends, complete strangers, family members, or co-workers. They can also be drawn from a mix of social networks, such as when one watches a video with a mix of friends and strangers.

Ultimately, collaborative online video sites can choose the type of audience to which to cater. By integrating with existing social networks (e.g., Facebook), they can create an environment in which friends or friends-of-friends watch videos together. Viewers in this environment would have greater feelings of trust toward other viewers, either because they already have pre-existing relationships or because the other viewers are shared acquaintances. Thus, this environment promotes the maintenance of existing social ties, as well as the creation of missing ties among the people in the friend-of-a-friend network.

Another option is to promote the formation of new ties across different social networks by encouraging strangers to watch together. One way to increase interpersonal attraction is to highlight how others are similar in preferences, attitudes, or values (Ren, Kraut, & Kiesler, 2007). In the collaborative watching case, a shared interest (or disinterest) in a particular video serves as a signal for similarity in preferences. Thus, collaborative online video sites can use features such as one's video ratings or video watching history to make the similarities among viewers visible, and hence encourage complete strangers to watch together.

In this dissertation, I examine groups with two types of pre-existing relationships: groups of people who know each other and consider each other as friends, and groups of people who are complete strangers and have not met each other prior to the studies. By examining both types of groups, I am able to understand how collaborative watching affects mutual feelings of liking and closeness for viewers with existing relationships (i.e., maintaining existing ties), and how it helps strangers get to know one another (i.e., creating new ties).

3.5. PLAYBACK MODEL

Online video sites differ in how they can present videos to their users. In the case of sites like Justin.TV and UStream.TV, viewers all watch the same video at the same time. This is a streaming model for watching online video, as viewers are watching a live video stream at the same time.

Another model of video playback is the playlist model. There are two variants of this model. Personal playlists, used by sites such as YouTube Streams and Gaia Online, allow each individual viewer to queue up their own personal set of videos to watch. In this case, viewers in the same chat group may be watching different videos from each other. Thus, since viewers chatting together choose which videos they want to watch and when they want to watch them, they are not necessarily synchronized with each other with respect to their content.

Group playlists are also possible, and are used in the Social Video application discussed in Chapter 12. In this case, the members of a chat group share an individual playlist for the entire group, and their video playback is synchronized to each other. Thus, viewers chatting together in this model are synchronized with respect to their content. The key differences between the streaming and playlist models are given in Table 3-3.

Table 3-3. Comparison between the playlist and streaming video models.

Model	Description	Example sites / systems
Playlist (personal)	Audience members watch different videos; interaction may be synchronous or asynchronous	YouTube Streams, Gaia Online, Hulu
Playlist (group)	Audience members watch together from a queue of videos; interactions are generally synchronous	Social Video (Chapter 12)
Streaming	Audience members watch the same video at the same time; interactions are generally synchronous	Justin.TV, UStream.TV, all P2P video streaming systems

These models have implications for social interaction as the amount of explicit common ground is different. For example, in the streaming model, viewers watch the same video together, and that video provides a shared source of conversational topics. Each viewer in the chat will know that the other viewers in the chat group will understand any deictic references they make to the video, such as “*that* is funny” or “isn’t *this* great.” In the personal playlist model, viewers may watch different videos from each other. References to or comments made about a video may not be understood by other viewers, as they may not have watched that particular video yet.

In Chapter 9, I address the question of how these models impact the sociability of the chatting and watching experience, and how conversations differ between groups of friends watching the same videos at the same time (mirroring a streaming model) and groups of friends watching the same set of videos, but in a different order from each other (mirroring a personal playlist model). We will see that both models are capable of supporting conversation amongst friends, although viewers in the playlist model talk less about the content of the videos they watch, and query each other more to understand the viewing state of their friends (e.g., which video they are currently watching).

3.6. SUMMARY AND CONCLUSIONS

- This chapter presents a framework for the design of collaborative online video experiences in terms of the video content they offer, the technology they use to distribute video data, the device on which viewers consume

content, the scope of interactions among viewers, and the model of video playback (streaming or playlist).

- This dissertation primarily focuses on the short, entertaining video clips commonly found on sites like YouTube. Full-length feature films are discussed briefly in Chapter 6, and content such as political debates and news broadcasts are discussed briefly in Chapter 17.
- Video distribution technology determines the size of the viewing audience and the cost to deliver video to that audience. As peer-to-peer technologies can support large audiences at little cost, the focus of Part III is on designing collaborative online video experiences for large audiences.
- “Online video” can be viewed on many devices other than computers, such as televisions and mobile phones. This dissertation focuses only on the case of viewing video on computers.
- Interactions among viewers can be synchronous or asynchronous, and they can occur in the same place or remotely. Viewers can have a variety of relationships to other viewers, such as being close friends or complete strangers. They can also interact with each other using textual, auditory, or visual media. This dissertation focuses on friends and strangers who chat with each other synchronously, using textual and/or auditory media.
- Viewers can watch the same videos at the same time (streaming model or group playlist model) or different videos from each other (personal playlist model). Both of these models are examined in the Text vs. Audio study (Chapter 9).

4.

SOCIAL ONLINE VIDEO: A SURVEY OF WATCHING VIDEOS ON YOUTUBE

At the outset of this dissertation, little research had been conducted to understand why people watched video online and the extent to which online video promoted social interactions around it. To better understand the social aspects of online video, I conducted a survey of YouTube users at Carnegie Mellon. At the time of the survey, in late 2007, YouTube was the most popular site for watching videos online (comScore, 2007). As of this writing in 2009, YouTube maintains its standing as the most popular online video site (comScore, 2009).

Prior research has shown the importance of sharing media as a way to get to know others better (Vaida et al., 2005), and for keeping current with television content in order to maintain common ground with others (Brown & Barkhuus, 2006). Thus, it seems that online video can be a socially incorporating experience. The Pew Internet and American Life Project study of online video found that 57% of online video viewers surveyed share links to videos they find with others (Madden, 2007). Another 57% reported watching online video with other people, such as friends or family. They report: "The picture of the lone internet user, buried in his or her computer, does not ring true with most who view online video." (Madden, 2007, p. iii). In this chapter, we will see that online video is often experienced socially, from sharing videos with others to chatting with them with over instant messaging while watching.

4.1. A SURVEY OF YOUTUBE USERS

In order to understand the usage patterns, motivations for usage, and benefits derived from online video, I conducted a survey of YouTube users at Carnegie Mellon. At the time of the survey, YouTube was the most popular online video site. Thus, asking questions about YouTube in particular enabled me to reach the widest possible audience of online video users.

4.1.1. SURVEY QUESTIONS

Table 4-1 summarizes the research questions addressed in this survey. They fall into five broad categories: basic usage patterns, the value and benefits viewers derive from watching videos online, the methods viewers use to locate videos to watch, the social behaviors around sharing and commenting on videos, and how personality translates to engaging in social behaviors.

Table 4-1. Summary of research questions for the survey of YouTube users.

Basic usage patterns	How often do people watch, what types of content do they watch, and from where do they watch?
Value and benefits	What value do people derive from watching videos online? For what reasons do they watch online?
Finding and sharing videos	How do people find content to watch? With whom do they share content? Do they generate and share their own content?
Social behaviors	Do people watch videos socially or in isolation? Do they comment on videos?
Personality	What types of people are attracted to watching videos online? Does one's personality predict whether they will engage in social behaviors?

4.1.2. RECRUITMENT

The survey was run for 17 days, from the end of November to the middle of December in 2007. Part of this time period corresponded to when students were preparing for and taking final exams. Therefore, there is a potential historical bias as students' habits, moods, and propensity to respond to the survey may have been

altered because of their exams. In Section 4.6, we search for evidence of this bias in our results and find none.

Survey respondents were recruited using several methods, including word of mouth, postings to a popular CMU newsgroup (misc.market), and fliers on campus. To motivate participation, several raffle prizes were offered for completed surveys: a grand prize of an iPod Nano (valued at \$150), or one of ten Starbucks gift cards (valued at \$5 each).

4.1.3. RESPONDENTS

A total of 301 responses were received from the survey. Of these, 23 (7.6%) were incomplete and removed from the sample. The final sample consists of 278 responses.

The mean age of respondents was 23 years (SD = 6.8 years). More men responded to the survey than women (62% vs. 38%). Sixty-two percent of respondents were undergraduates, 28% were graduate students, and 9% were faculty or staff.

Carnegie Mellon has a unique organizational structure, with different academic disciplines federated across seven different colleges. To determine the role of respondents within the university, respondents were asked to which college they belonged. A comparison of these responses with the demographics of the general student body at the time the survey was conducted is shown in Table 4-2. Enrollment data come from the CMU Quick Facts report (Carnegie Mellon Institutional Research and Analysis, 2007).

Table 4-2. Demographics of survey respondents and the general student population. (*) Faculty and staff responses have been excluded to compare with enrollment data.

College	Sample* (%)	Enrollment (%)
Science & technology		
Carnegie Institute of Technology (CIT)	29.7	25.3
Mellon College of Science (MCS)	12.4	9.6
School of Computer Science (SCS)	20.5	12.4
Humanities & arts		
College of Humanities & Social Sciences (H&SS)	16.9	13.4

College	Sample* (%)	Enrollment (%)
College of Fine Arts (CFA)	5.2	11.5
Business		
Tepper School of Business (Tepper)	10.0	14.2
Heinz School of Public Policy & Mgmt. (Heinz)	4.0	6.1
Other		
Interdisciplinary programs (e.g., BHA)	1.2	7.4

Overall, the sample was biased toward technical and scientific colleges, with 62.6% of student respondents from CIT, MCS, and SCS, versus 47.3% enrolled in those colleges. The sample was also biased toward undergraduates, and consisted of 62% undergraduate and 28% graduate students, compared to the student population of 55% undergraduate and 45% graduate students.

With regard to the ethnicity, respondents were evenly split between White/Caucasian (46%) and Asian (49%, including Indian and Chinese); 5% reported other ethnicities (including Black, Hispanic/Latino, Mixed Race, and Native American), and four respondents did not report any ethnicity. Seventy percent of respondents reported being native English speakers. Non-native speakers were more likely to be Asian ($\chi^2 (2, N=274) = 71.0, p < .001$) and graduate students ($\chi^2 (2, N=275) = 40.3, p < .001$).

4.2. BASIC USAGE PATTERNS

The first questions on the survey asked about how often respondents visited YouTube, where they watched, what they watched, and how much they enjoyed using the site.

In order to guard against incorrect categorizations of general behavior, respondents were asked how often they visited YouTube in the past two weeks. About 27% reported visiting daily or more than once per day ("heavy users"), and 73% reported visiting several times or once or twice in the past two weeks ("occasional users"). Respondents were also asked about how many videos they had watched 'yesterday', again to guard against a memory bias. Overall, respondents reported watching 2.1 (SD = .93) videos on average. As expected, heavy users reported watching more

videos than occasional users (heavy: $M [SD] = 2.92 [.82]$ videos, occasional: $M [SD] = 1.77 [.77]$ videos, $F [1,276] = 115.5, p < .001$).

Respondents were next asked where they watched YouTube videos: at home, in the office, in class or in the computer clusters on campus. The scale of “never,” “sometimes,” and “often” was used to gauge frequency. An “N/A” option was also included as not all of the specified locations applied to everyone in the sample (undergraduates, for example, generally do not have offices). Overall, most respondents reported watching at home (97% reported “often” or “sometimes”), and a significant number reported watching “often” or “sometimes” in the office or in the computer cluster (42% in both cases; 16% reported N/A for offices and 3% reported N/A for the computer cluster). Finally, 16% (43 people) reported watching videos during class, with only 3% reporting N/A, suggesting that most respondents had the opportunity to watch in class.

Our question about location also included an open-ended response field for other places where people watch YouTube videos, and respondents reported several interesting locations: at friends’ rooms or homes (7 people), on their laptop (4 people), in a graduate lounge or studio (3 people), in coffee shops (2 people), and on their iPhone or iPod Touch (3 people). All of these locations share the characteristic that people are mobile when visiting them. These results show that some people watch videos while they are in a location outside the boundaries of home or work.

4.3. VALUE AND BENEFITS

In order to assess how much respondents enjoy using YouTube, I constructed a four-item enjoyment scale, shown in Table 4-3. Answers were on a 5-point Likert scale, corresponding to “strongly disagree,” “disagree,” “neither agree nor disagree,” “agree,” and “strongly agree.” Cronbach’s alpha for the four-item scale was .71, although by excluding item 2, alpha increased to .77. Thus, the final enjoyment scale used only items 1, 3, and 4.

Table 4-3. YouTube enjoyment scale. (*) Item 2 was dropped in the analysis to increase reliability.

#	Item
1	I enjoy using YouTube
2*	YouTube is an important part of my life
3	Watching YouTube is fun
4	I usually find interesting videos on YouTube

Overall, respondents enjoy using YouTube (M [SD] = 4.04 [.60]). Heavy users enjoy it more than occasional users (heavy: M [SD] = 4.25 [.64], occasional: M [SD] = 3.97 [.57], $F [1,276] = 12.7, p < .001$). This finding is consistent with the observation that enjoyment is correlated with how often respondents visit YouTube ($r = .22, p < .001$).

To assess the benefits respondents receive by watching videos, I created a list of reasons why people might watch YouTube. This list was based in part on the list of benefits of online communities used in the HomeNet study (Kraut, 1995). This list was comprised of 29 items, although 17 of the items did not receive enough support from respondents to warrant inclusion in the analysis (items with less than 20% of respondents choosing them were dropped). This left 12 reported reasons for using YouTube. Examples of items excluded from this analysis were to “get the news” (7.9%), to “overcome loneliness” (9.7%), to “get information related to my finances or how to invest my money” (1.1%), and to “meet new people” (0.7%).

A discriminatory factor analysis of the 12 reported reasons showed that they loaded on five factors. After examining the factors each item loaded on, the factors were labeled Activity, Feelings, Entertainment / Knowledge, Access, and Misc. They are listed in Table 4-4. Note that only one item, “discuss with others,” relates to watching YouTube as a social activity. All of the other reported reasons are “selfish,” in that the value derived is solely for the benefit of one’s self. This is in contrast to the wide range of social practices in which respondents reported engaging, such as sharing videos or watching together, which is discussed further in this section. Thus, although the value people derive from watching videos seems to primarily be individualistic, the prevalence of social practices around watching video suggest that people are able to derive individual value while watching in a social situation.

Table 4-4. Reasons for using YouTube. Percentages of respondents who affirmed each reason are listed in parentheses.

Activity	Access
Kill time (73%)	Watch things I can't get on television (53%)
Procrastinate (68%)	Watch clips of shows I missed on television (43%)
Overcome boredom (62%)	
Feelings	Misc.
Elevate mood (42%)	Listen to music (62%)
Release tension (35%)	Discuss with others (29%)
Entertainment / Knowledge	
Entertainment (83%)	
Educate myself (29%)	
Get information about a hobby (25%)	

4.4. FINDING AND SHARING VIDEOS

There are many ways one can find videos to watch on YouTube. For example, they can be recommended by others, linked from web pages, featured on YouTube's front page, or found by browsing video tags or related videos. Respondents were asked how often they perform these activities when finding videos to watch, using the frequency scale of "never," "rarely," "sometimes," "often," and "almost always." For the purposes of this analysis, "almost always" and "often" will be referred to as "frequently."

The methods for finding videos on YouTube can be broken down into three classes: social recommendations, for when another person recommends a video; system recommendations, for when YouTube recommends a video to you; and self-recommendations, for when you actively browse or search for a video. A breakdown of the frequency of use for each of these methods is shown in Table 4-5.

Table 4-5. Frequency of usage for different types of video recommendations. “Frequent” is the sum of “Often” and “Almost always” responses. Reported numbers are the percentage of respondents in each category.

Recommendation type	Never	Rarely	Sometimes	Often	Almost always	Frequent
Social						
Recommended by friends	2	9	36	41	12	53
Linked or embedded in web pages	6	16	37	37	4	41
System						
Featured on the front page	35	36	21	7	1	8
Browsing related videos	4	9	38	40	9	49
Self						
Using YouTube’s search feature	4	12	27	36	21	57
Browsing video tags	38	31	23	7	1	8

Social recommendations are quite popular, with about half of respondents reporting that they frequently received recommendations from their friends, and about 41% frequently watching videos linked or embedded in web pages. The latter method is considered “social” because, even though a social interaction may not have taken place to make the recommendation, the video was chosen by a person to be seen by others (as opposed to being chosen by a machine). Respondents paid just as much attention to YouTube’s recommendations as to social recommendations, with 49% reporting that they browse related videos. Counter to our expectations, the videos featured on YouTube’s front page were not as popular, with only 8% reporting that they frequently looked to the front page to find videos to watch. Finally, self-recommendations were also important, but only for searching for videos (57%) and not browsing them (8%).

The questions about sharing videos made a distinction between sharing with individuals and sharing with groups, such as clubs or social networking groups. Respondents were asked with whom they shared videos (Table 4-6), and the means they used to share them (Table 4-7).

Table 4-6. People with whom videos are shared. “Frequent” is the sum of “Often” and “Almost always” responses. Reported numbers are the percentage of respondents in each category.

Entities	Never	Rarely	Sometimes	Often	Almost always	Frequent
Individuals						
Friends	3	4	26	39	28	67
Family members	22	27	31	15	5	20
Co-workers or classmates	20	21	34	22	3	25
Groups						
Clubs or organizations	60	18	16	6	0	6
Social networking groups	61	17	18	4	0	4

Table 4-7. Methods used for sharing videos. “Frequent” is the sum of “Often” and “Almost always” responses. Reported numbers are the percentage of respondents in each category.

Sharing method	Never	Rarely	Sometimes	Often	Almost always	Frequent
Sharing to individuals						
Email (to an individual)	31	22	31	13	3	16
Instant messaging	27	14	32	19	8	27
Show in person	11	15	41	27	6	33
Sharing to groups						
Email (to a mailing list)	68	17	10	4	1	5
On my web site or blog	70	15	11	4	< 1	4
On a social networking site (e.g., Facebook, MySpace)	54	22	17	5	2	7

Overwhelmingly, sharing among individuals was more popular than sharing with groups. Respondents reported a greater frequency of sharing videos with friends (67%), family members (20%) and co-workers or classmates (25%) than with clubs and organizations (6%) or social networking groups (4%). Further, although email (16%) and instant messaging (27%) were the most popular means of sharing videos online, showing videos in person (33%) was the most popular method reported for sharing videos.

In any online community, the value of the community is directly proportional to the quality of its content. Without new and interesting content, a community will

struggle to recruit new members and retain existing ones (Butler, 1999; Butler et al., 2002; Lee et al., 2009; Nonnecke & Preece, 2000). In the case of YouTube, this requirement equates to a continuous need for members to upload new and entertaining videos. Thus, to understand the extent to which YouTube viewers uploaded content, respondents were asked if they uploaded their own videos to YouTube. Only 37 people (13%) reported uploading videos. To understand what types of users upload videos, we perform a logistic regression. The explanatory variables in this regression are college, ethnicity, gender, age, being a native speaker, position within the university (student, faculty/staff, etc.), and whether the user is a heavy user. Being a heavy user is the only significant predictor of whether a user uploads a video ($\chi^2 = 3.69$, $p = .05$). A contingency analysis shows that heavy users are more likely to upload videos than occasional users (heavy users who upload: 21%, occasional users who upload: 10%, $\chi^2 (1, N=278) = 5.3$, $p = .02$).

Respondents who reported uploading videos were asked about why they uploaded videos. These reasons are shown in Table 4-8. Again, we make the distinction between “selfish” motivations, “social” motivations, and “mixed” motivations (as they can go either way), and find that the most popular reasons for uploading fall into both categories.

Table 4-8. Motivations for uploading videos to YouTube. Percentages are of N = 37 people who reported uploading videos.

Motivation	%
Selfish	
Because I can	49
It's fun	49
To store my videos	38
Social	
Share with others	73
Get feedback from others	30
Responding to a video	6
Mixed	
Exhibit my creativity	43
Made a good video	16
Engage in an important issue	6

4.5. SOCIAL BEHAVIORS

Respondents were asked about a number of social and non-social activities in which they might have engaged while watching videos, including email, instant messaging, talking on the phone, and studying. A breakdown of the percentage of respondents that reported frequently engaging in these activities is given in Table 4-9.

Table 4-9. Activities performed while watching videos. “Frequent” is the sum of “Often” and “Almost always” responses. Reported numbers are the percentage of respondents in each category.

Activity	Never	Rarely	Sometimes	Often	Almost always	Frequent
Social						
Instant messaging	20	11	27	31	11	42
Email	14	16	41	23	6	29
Talk to someone in person	23	18	36	21	2	23
Web forums	57	16	18	7	2	9
Talk on the phone	37	28	26	8	1	9
Talk to someone using voice or video conferencing (e.g., Skype)	66	18	11	5	<1	5
Chat rooms (e.g., IRC)	75	14	7	3	1	4
Non-social						
Eating	10	15	50	23	2	25
Homework / studying	28	18	38	14	2	16

The most frequent activities were instant messaging (42%), email (29%), and talking to someone in person (23%), suggesting that socializing while watching is quite popular. Respondents also reported engaging in non-social activities such as eating (25%) and doing homework or studying (16%).

Interesting to note are the respondents who reported frequently watching videos while doing homework or studying. These respondents remind us of the 43 respondents who reported watching videos in class. However, of the 46 respondents who frequently watch while studying, only 15 (33%) reported watching during class; thus the two groups are similar, but not identical.

Finally, one approximation to learning how often participants want to focus solely on the video they are watching without distraction (i.e., from email or IM, not from the

real world) is to know how often they make the video full screen. By watching in full screen, the only display on their screen is the video; other programs such as instant messaging and email are hidden. Twenty-four percent of participants reported frequently using full-screen mode, with 5% reporting “almost always,” and 19% reporting “often.” This finding suggests that, for these participants, there are times where they do wish to immerse themselves in the video and not be bothered by other events or activities.

4.6. PERSONALITY

Three personality constructs were measured in this survey in order to determine whether different types of people have different usage patterns. These constructs were extraversion and openness from the Big-5 inventory (Goldberg et al., 2006) and depression using the CES-D Depression Scale (Radloff, 1977). These constructs were measured on 5-point Likert scales, with 5 representing the highest level of the construct. A median split was used to classify each respondent as more-extraverted/less-extraverted, more-open/less-open, and more-depressed/less-depressed. Although the latter classification may carry negative social connotations, this effect is not the intention; more/less is merely used to label respondents based on the outcome of the median split.

Descriptive statistics of extraversion, openness, and depression are shown in Table 4-10. Overall, extraversion and openness scores were centered on the middle of the scale. Mean depression scores were unexpectedly low given that our survey was administered right before and during final exams ($M [SD] = 1.9 [.47]$ of 5). Therefore, we do not find evidence for a historical bias in our results in terms of the mood of our respondents.

Table 4-10. Distributions of personality constructs. Responses are on a 5-point scale, with 5 representing the largest value for each construct.

Personality construct	Mean	Stdev.	Median
Extraversion	3.1	.62	3.1
Openness	3.7	.52	3.7
Depression	1.9	.47	1.8

We first examine the role personality plays on YouTube usage. Using each personality construct, we predict how often respondents reported visiting YouTube.

A logistic regression using extraversion, openness, and depression to predict frequency of visitation shows that depression is a significant predictor of usage ($\chi^2 = 10.7$, $p = .03$). A contingency analysis shows that more-depressed people are more likely to visit YouTube daily or more often (34%) compared to less-depressed people (19%). There is also a weak but significant correlation between depression and the number of videos respondents reported watching in the day prior to taking the survey ($r = .13$, $p = .02$). Although this analysis does not establish a causal link between watching YouTube and feelings of depression, it does suggest a relationship between feelings of depression and frequency and amount of usage.

Next, we examine how personality predicts social behaviors while watching online video. We consider the most popular social and non-social behaviors while watching video: checking email, instant messaging, talking to someone in person, eating, and doing homework or studying. Reported levels of other behaviors, such as using chat rooms or web forums, were quite low and do not show any significant differences among groups. For this analysis, behaviors were encoded into a binary variable: 1 if the respondent reported engaging in that behavior “rarely,” “sometimes,” “often,” or “almost always,” and 0 if the respondent reported “never”..

Email use. A binary logistic regression using extraversion, openness, and depression to predict email usage shows that extraversion tends to predict email usage ($\chi^2 = 3.0$, $p = .08$). A contingency analysis shows that 89% of more-extraverted people email while watching, versus 83% of less-extraverted people email while watching.

Instant messaging. Depression significantly predicted IM use while watching ($\chi^2 = 8.3$, $p < .01$). A contingency analysis shows that 86% of more depressed people talk to others online while watching, compared to 73% of less-depressed people.

Talking in person. Extraversion significantly predicted whether one talks to others in person while watching YouTube videos ($\chi^2 = 6.0$, $p = .01$). A contingency analysis shows that 79% of more-extraverted people talk to others in person while watching, compared to 75% of less-extraverted people.

Eating. Extraversion tended to predict whether one eats while watching YouTube videos ($\chi^2 = 3.4$, $p = .06$). A contingency analysis shows that 88% of more-extraverted people eat while watching, compared to 92% of less-extraverted people.

Homework/studying. Extraversion and openness tended to predict whether one does homework or studies while watching YouTube videos (extraversion: $\chi^2 = 3.7$, $p = .06$; openness: $\chi^2 = 3.5$, $p = .06$). Depression was a significant predictor of doing homework or studying while watching ($\chi^2 = 9.6$, $p = .002$). For extraversion, a contingency analysis does not show much difference in homework/studying habits: 72% of more-extraverted people reported doing homework while watching, versus 73% of less-extraverted people. For openness, a contingency analysis shows that 69% of more-open people do homework or study while watching, compared to 76% of less-open people. Finally, a contingency analysis for depression shows that 76% of more-depressed people do homework or study while watching, compared to 68% of less-depressed people.

4.6.1. DISCUSSION

YouTube is quite popular among CMU students, as well as some faculty and staff members. Almost a third of our sample reported visiting YouTube daily or more than once per day, with the other two thirds visiting once, twice, or several times a week. As expected, almost all respondents reported watching from home (97%). However, significant numbers of respondents reported watching videos “sometimes” or “often” in the office or in the computer cluster (42% in both cases). This finding is interesting because it suggests that the culture of the office environment and the computer cluster (perhaps the closest approximation undergraduates have to an office) accepts watching videos; if watching videos in the office or cluster is seen as an inappropriate activity, we would expect fewer people to report this behavior.

Also of note are the 43 people who reported watching videos in class. Although the dynamics of the home, office, and computer cluster environments allow for people to take breaks and distract themselves by watching a video, the classroom is a place where it is typically unacceptable to engage in outside activities. However, this norm has not historically prevented students from goofing off during class. Online video now joins the cadre of napping, note passing, and web surfing as classroom distractors. Although it is unclear whether watching videos in class was done in solitary or in the company of others (perhaps in the back row, or more acceptably, as part of the lecture), it does seem to have a higher potential for embarrassment than passing notes or surfing the web. As most videos are accompanied by sound, students who forget to mute the volume of their laptop, plug in headphones, or

prevent themselves from laughing out loud may be disruptive to the class. For these 43 students, either this potential for causing a disruption seems to not be enough motivation to not watch during class, or they reported watching videos that were part of their lecture. Either way, it seems that online video is moving into the classroom, for better or for worse.

Respondents reported using YouTube for a variety of reasons. The most popular was for entertainment (83%), followed closely by killing time (73%), procrastinating (68%), and overcoming boredom (62%). This distribution paints a somewhat bleak picture of YouTube as a place to go for wasteful leisure. However, significant numbers of respondents use YouTube to better themselves or their emotional states by learning about a hobby (25%) or educating themselves (29%), or by elevating their mood (42%) or releasing tension (35%). Thus, YouTube does seem to provide emotional and educational outlets for those who seek them.

Finding content on YouTube amounts to receiving recommendations for that content socially, directly from YouTube, or by searching or browsing. All three general methods were popular among respondents, although the popularity of specific means for finding content varied. For example, browsing the videos featured on the YouTube front page was not very popular, with only 8% of respondents reporting that they frequently found videos this way. This finding seems counter-intuitive, as home pages are often used to feature prominent or important content; yet, we see much more popularity in browsing videos related to the one just watched (49% do this frequently). Although we do not know for certain why respondents do not use the front-page recommendations, there are several likely explanations. First, respondents may not need to resort to finding videos on the front page, simply because other methods of finding videos are so successful; thus, they would not need to search for additional content. Or, respondents may simply trust the set of related videos more than the set of featured videos because they know they enjoyed the current video and want to watch something similar. In this case, system recommendations would be most useful when they are relevant to the current context (i.e., the current video being watched, or perhaps the current conversation about a video), and not when the recommendations are based on mass aggregated behavior (i.e., a video that everyone is watching).

Another unpopular method for finding videos was by browsing the tags applied to videos. Only 8% of respondents reported doing this frequently, compared to 57%

who reported frequently using YouTube's search feature to find videos. Typically, browsing is used for open-ended information-seeking tasks, and searching is used for focused information needs (McDonald & Chen, 2006). The popularity of searching suggests that viewers typically have focused interests when watching videos. Further, the unpopularity of using tags to browse for videos calls into question the usefulness of tags: if people don't use tags, why have them at all? One answer is that tags can be used to find related videos. Thus, applying tags to videos increases peoples' likelihood that related videos are really related. Another answer is that tags can improve the relevance of videos found through the search feature. Therefore, although tags do not seem useful for browsing videos, they can still be useful for helping people find videos.

As for sharing videos, respondents reported frequently sharing with individuals, such as friends (67%) and family (20%). Less popular was sharing with groups, such as sharing videos with mailing lists (5%), with social networking groups (7%), or on web sites (4%). One explanation for this finding is that sharing videos with individuals is relatively easy using instant messaging or email. It is also a personal experience. For example, the thought process of sharing a video with a friend may be akin to, "I am sharing this video because I think she will like it" or "I am sharing because I liked it." Sharing videos with a group may be less desirable for people because it is less personal, or because it is simply more difficult to do (e.g., people may have had difficulties in publishing or sharing links online). Although sharing videos with groups was not popularly reported in this survey, it is possible that technological advances since the survey was run have made it easier for people to share videos with groups, and thus increased its popularity. For example, sites including Facebook, MySpace, digg, and reddit make it easy to share video links with large groups of people online. Indeed, 87% of our respondents reported having a profile on a social networking site, giving them easy access to a video sharing feature. Therefore, as the technological hurdles to sharing videos with groups online are decreasing, more attention can be paid to the new, social video applications being created for groups of people online.

Sharing videos by showing them in person was quite popular. A third of respondents reported frequently showing videos to others in person. Only 11% of respondents (31 people) reporting never showing videos in person. These results support the argument that, for many people, online video is experienced socially. Further, the most popular reason for uploading videos to YouTube was "to share with

others" (73%), demonstrating that sharing is a significant motivation for creating and uploading videos.

Socializing around online video does not stop with the act of sharing videos. Respondents also reported frequently engaging in a number of social behaviors while watching videos, including instant messaging (42%), email (29%), and talking to someone in person (23%). Although we do not know if respondents' emails or instant messages were related to the videos they watched, we do see evidence that they are not passively watching videos online. Rather, the picture emerges of viewers who actively multitask between watching videos and other activities such as email, instant messaging, or even eating (a quarter of respondents reported frequently eating while watching). These findings suggest that people will be receptive to interaction features in an online video interface since they are already used to multitasking while watching videos. For those users who already engage in self-distraction while watching, chatting while watching may not be additionally distracting. This is a promising finding because it supports the idea of combining interaction with video. The issue of distraction is examined in closer detail in Chapters 7-9.

Other methods for socializing in person and online were not as popular. Only 4% of respondents reported frequently using chat rooms while watching, and 25% reported using chat rooms while watching at all. Nine percent reported frequently using web forums while watching, and 43% reported using web forums while watching at all. Five percent reported frequently using voice chat while watching, and 34% reported using voice chat while watching at all. Thus, many respondents have at least tried these types of computer-mediated communications media while watching. Although we do not know why these media are used less frequently than email and instant messaging, there are several possibilities. One is simply that the base rates of usage are low; everyone at CMU has an email account, and many use instant messaging, but it is unknown how many visit chat rooms or post to web forums. Another explanation is that group communication is more difficult than one-to-one communication while watching a video, perhaps because there is more content to read. Voice conferencing may also be difficult while watching a video, as the audio channel of the chat interferes with the audio channel of the video. These hypotheses about chatting in groups and audio interference are examined in the laboratory studies discussed in Part II, and the results suggest that neither leads to a poor experience. Thus, the most likely explanation is that there is simply a low base

rate of usage for chat rooms, web forums, and voice chat, and not that these media inherently do not work well with video.

Finally, we found that personality constructs such as extraversion, openness, and depression were able to predict how often respondents visited YouTube, as well as the social and non-social behaviors in which they engaged while watching. Depression was linked with higher visitation rates, and it was correlated with the number of videos respondents reported watching in the day prior to taking the survey. However, this analysis cannot establish causality, so it is unclear if people who are more depressed are more likely to visit YouTube, or if visiting YouTube more often results in higher feelings of depression.

We also found that extraversion tended to predict whether or not respondents would email others while watching videos, and it significantly predicted whether respondents talked to other people in person while watching. These findings intuitively make sense, as higher levels of extraversion are typically associated with higher levels of social activity. However, as extraversion did not predict instant message usage while watching, there may be other personality constructs that determine social behaviors while watching. For non-social behaviors, we found that less-extraverted people were more likely to eat while watching. This finding is congruent with how we expect less-extraverted people to behave, by engaging in a solitary activity while watching rather than a social one. Finally, we found that more-depressed people were more likely to do homework or study while watching, which may be reflective of a correlation between needing to study for final exams and feeling depressed about the exams.

Overall, the results of this survey show that online video is experienced socially, and that users are likely to accept and use features that let them interact with each other while watching. In Part II, I discuss results from experimental studies that examine the usage of such interaction features, their distractive effects, and their impact on sociability.

4.7. SUMMARY AND CONCLUSIONS

- Many people reported watching YouTube videos to be entertained, to kill time, to procrastinate, and to overcome boredom. Some reported watching them to learn about a hobby, to educate themselves, to elevate their mood, or to release tension.
- Videos enable social interactions. Sharing videos with friends, family, or co-workers was common, as was socializing through IM, email, or in person while watching. Sharing with others was also cited as a common reason for uploading videos to YouTube. These findings motivate the need for further, controlled research on the social interactions that occur as viewers watch videos together. This research is discussed in Part II of this dissertation.
- One's propensity to socialize around videos was related to their personality traits. Extraversion was associated with higher levels of social activity around videos (e.g., emailing), and introversion was associated with higher levels of non-social activity around videos (e.g., eating). Depression was associated with higher levels of YouTube usage. These findings may account for individual differences in the enjoyment of collaborative watching.

Part II: Simultaneous Watching and Chatting

Part II presents a series of empirical studies that examine the collaborative online video watching experience from the perspective of a small group of viewers. These studies quantify the extent to which viewers are distracted, the effect of chatting on their feelings of closeness to and liking of each other, and how different chat media – text and voice – are experienced.

5.

EMPIRICAL STUDIES OF CHATTING WHILE WATCHING

The core activities in a collaborative online video site are watching videos and chatting with others. Chapter 1 discussed the combination of these two activities from two seemingly dichotomous viewpoints. According to the human factors literature, chatting while watching video will not be desirable because both activities require one's attention. Chat distracts a viewer from watching the video, and watching the video distracts a viewer from chatting. Thus, chat can interfere with a viewer looking for an entertainment experience, and the video can interfere with a viewer looking for a social experience.

Opposing this viewpoint is that of the social capital literature. Although it says nothing of distraction, it does make a compelling argument for the combination of chat with video. By incorporating social interaction with an activity that, in some circumstances, promotes isolation, we can create opportunities for people to improve the quality of their social networks. Thus, online video experiences ought to be collaborative, and distraction (if present and detrimental to the experience) should simply be mitigated and minimized to the fullest extent possible. In other words, the social capital viewpoint argues that the positive social consequences of collaborative watching outweigh any potential negative effects from viewers being distracted.

Both of these arguments are compelling⁴, but they merely state opposite ends of the spectrum: either do not mix both activities because they will overwhelm the user (distraction argument), or mix them because they may have positive social consequence (sociability argument).

The goal of the chapters in Part II is to determine the extent to which each argument is correct. Is it distracting to chat with others while watching a video? Can chatting with others lead to positive social outcomes? I address these questions by quantifiably measuring the extent to which chatting while watching is distracting and the extent to which social relationships are impacted by watching with others.

To address these questions, I ran four empirical research studies that examined the interplay between interaction and distraction. These studies employed several experimental designs, and the research methods included both laboratory and field work. Table 5-1 gives a summary of the studies and their goals.

Table 5-1. Summary of the empirical studies of collaborative online video watching.

Study	Method	Venue	Goals	Chapter
MovieLens	True experiment	Field	Preliminary study of watching and chatting	Chapter 6
Chat Distraction	True experiment	Laboratory (simulated)	Quantify distraction	Chapter 7
Cartoon	Quasi-experiment	Laboratory (live)	Reduce distraction, quantify social effects	Chapter 8
Text vs. Audio	True experiment	Laboratory (Live)	Compare chat media on sociability and distraction	Chapter 9

Note that three of these studies – MovieLens, Chat Distraction, and Text vs. Audio – were true experiments as the assignment of participants to the experimental conditions was entirely random. The Cartoon study was a quasi-experiment as pre-existing groups of friends and groups of strangers were used; participants could not be randomly assigned to be strangers or friends with the other participants in their group.

⁴ They are compelling in the sense that, before collaborative online video sites became popular, there was a general uncertainty about whether adding chat to video was a good idea. Since then, it has become clear that it was.

5.1. GENERAL METHODS

The empirical studies discussed in Part II use one of two general methods. They were either conducted in a real-world setting, in which people watched videos over the Internet from their home computers, or they were conducted in the laboratory, in which participants watched videos in a controlled setting. Further, the laboratory studies were either simulated, in which individual participants watched videos and read a pre-created chat transcript from a “wizard of Oz” chat group, or they were live, in which small groups of participants watched together. The strengths and weaknesses of each of these approaches are discussed below.

5.1.1. FIELD STUDIES

Field studies allow us to understand collaborative online video watching in a real-world context. They have the nice property of preserving the realism of the activity. Participation in a field study offers a semi-realistic setting for watching videos and chatting with other people over the Internet, using one’s own computer. This realism boosts our confidence that our results have external validity. External validity is important as it gives us confidence that our results will generalize to other real-world settings and populations. However, this ability to generalize our findings comes at the cost of being unable to observe interesting phenomena in isolation. For example, if we were to measure distraction in a field study, we must be aware that there are other sources of distraction than just the video: people may check their email, eat food, answer the phone, or walk out of the room, all while claiming to be participating in the study. Therefore, we turn to the laboratory to provide controlled conditions that eliminate these confounding variables.

5.1.2. LABORATORY STUDIES

Laboratory studies are used to carefully control the confounding factors that make it difficult to establish a causal relationship between manipulations and outcomes. They do this by placing participants in an environment in which their sole focus is to participate in the experiment, and the only information they have about their situation is given to them by the experimenter. This level of control allows the experimenter to compare the behaviors of multiple participants across similar situations, in which the only differences between those situations are those that

were manipulated by the experimenter. Control comes at a loss of generality, however, as participants may behave in a manner different from how they behave in their daily lives. For example, not all participants in our laboratory studies were frequent users of online video sites. Thus, although they participated in a collaborative online video experience in the laboratory, it was not necessarily a behavior they exhibited in their daily lives.

By combining both approaches – field studies and laboratory studies – we are more fully able to understand the activity of collaborative online video watching.

5.2. MEASURES

The studies in this chapter are primarily concerned with measuring three constructs: distraction, enjoyment, and sociability. The Text vs. Audio study measured several secondary constructs, including engagement and media preference. Secondary constructs (used only in the Text vs. Audio study) include mood and engagement. Other measures typically include demographics, online video watching behaviors, and questions about the specific manipulations experienced in the study. A full list of the surveys and scales used across the empirical studies can be found in Appendix B.

5.3. DISTRACTION

Distraction is the extent to which a viewer must manage their attention between the video and chat. When a viewer ignores the chat because their attention is focused on the video, we say that the video distracts the viewer from the chat. Likewise, when a viewer ignores the video because their attention is focused on the chat, we say that the chat distracts the viewer from the video.

I use two measures of distraction in this dissertation. The first is a self-reported measure that asks participants to rate how distracted they felt on an open 7-point scale anchored by “Not distracted at all” to “Very distracted”. The second measure of distraction quantifies how much attention participants paid to the video and the chat by asking memory recall questions about what they read in the chat and what they saw in the video. One caveat to using a memory measures for chat recall is that it can only be used in situations in which the chat transcript is known ahead of time

(i.e., in a simulated environment). Thus, chat memory is only measured in the Chat Distraction study.

5.4. ENJOYMENT

Enjoyment is the extent to which a viewer had fun participating in the study, watching the videos, and chatting with others. Study enjoyment was measured by asking participants to rate statements such as “I had fun watching the videos” and “The videos were entertaining.” Video enjoyment was measured by having participants rate the videos they watched (discussed further in Section 5.6.2). Chat enjoyment was measured by asking participants to rate statements such as “I enjoyed chatting with other people.” A full list of the enjoyment scales used in the studies is given in Appendix B.

5.5. SOCIABILITY

Sociability is the extent to which participants felt each others’ presence while watching together. Several measures of sociability are used in this dissertation. The first was a behavioral measure of simply how much participants chatted; if participants did not use the chat feature, or did not use it much, then the experience was not very social. Conversely, if participants used the chat feature quite a lot, the experience may have been highly social.

Self-reported measures of sociability were used as well. Participants were asked to rate their level of agreement with statements such as “I enjoyed talking with the people in my group while watching the videos” and “I would have preferred to watch the videos alone” to determine the extent to which they enjoyed having other people with whom to chat while watching the videos.

Participants were asked questions about their feelings toward the other participants in their group. The first set comprised a scale that measured liking – the degree to which participants like the other people in their chat group. For this scale, participants rated statements such as “They were friendly” and “I liked them.” A full list of the items on the liking scale is given in Appendix B.

Finally, the Inclusion of Other in Self scale (Aron et al., 1991) was used to measure momentary feelings of closeness between participants in a group. This scale

presents a series of progressively overlapping circles, one representing “Self” and the other representing “Other.” The amount of overlap between these circles determines the extent to which the participant feels close to the other participants. This scale was used in a pairwise fashion – each participant rated his or her feelings of closeness to each other participant in the group. An example of this scale is given in Appendix B.

5.6. VIDEOS

Each of the studies required a set of videos for participants to watch. In the MovieLens study, we showed a series of five feature films to participants. These films ranged in duration from one to two hours, and included films from different genres to appeal to a wide audience.

The Chat Distraction, Cartoon, and Text vs. Audio studies used shorter video clips. To be representative of the types of videos people watch on popular online video sites like YouTube, these videos were selected to meet the following criteria:

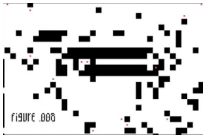
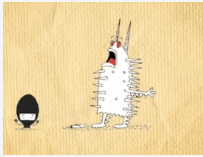

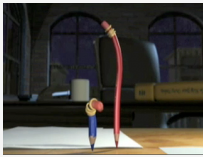




- Short in duration (3-7 minutes long)
- Minimally offensive (no strong language)
- Generally accessible to a broad audience



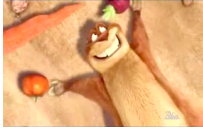




In addition, the Text vs. Audio study required videos that fit one additional criteria:

- Varied presence of verbal content (verbal content present vs. absent)

Table 5-2 shows a breakdown of each of the videos used in the Chat Distraction, Cartoon, and Text vs. Audio studies. Note that the videos used in the Chat Distraction study were identical to those used in the Text vs. Audio study; chronologically, the Text vs. Audio study was run before the Chat Distraction study, and thus the videos were recycled. In addition, the Tag Evaluation study (Chapter 14) used a subset of the Cartoon and Text vs. Audio study videos, and the Best Parts study (Chapter 16) used all videos from both studies.

Table 5-2. List of videos used in each study. Cartoon study videos came from Channel Frederator, a cartoon video podcast. Text vs. Audio (and Chat Distraction) study videos came from YouTube. The runtime of the video and a brief description of the video's genre and content are given.

Thumbnail	Video	Runtime	Description
Cartoon study videos			
	Emerge	3:53	Cartoon animation; comical introduction to cellular automata
	Fuggy Fuggy	4:53	Cartoon animation; humorous caricature of a ninja in training
	In the Rough	4:47	Cartoon animation; humorous tale of the plight of a prehistoric caveman
	Pen Pals	5:07	Cartoon animation; a pen and a pencil compete for the affections of a female pencil
	Penguin's Christmas	3:21	Cartoon animation; a sad penguin is cheered up by his friends on Christmas
	Plumber	6:04	Cartoon animation; a humorous tale of a plumber trying to fix a leak
	War Photographer	4:05	Cartoon animation; a battle-of-the-bands between rival Viking ships
Text vs. Audio study videos (also used in the Chat Distraction study)			
	Ali G - War	5:59	TV show clip; Ali G interviews General Brent Scowcroft (former National Security Advisor) about war, tactics, and strategy <i>verbal content</i>

Thumbnail	Video	Runtime	Description
	Brothas From the Same Motha	4:01	Cartoon animation; narrator makes a humorous case that the Bic pen is a close relative of Marvin the Martian <i>verbal content</i>
	Daughters	3:35	Live action skit; a secret agent must rescue his daughters from terrorists <i>verbal content</i>
	Gopher Broke	4:17	Cartoon animation; a gopher makes several thwarted but humorous attempts at rescuing stray vegetables from passing trucks <i>no verbal content</i>
	Paddy the Pelican	5:13	Cartoon animation; two bears stranded on an island receive help from a pelican in escaping from a mean sailor <i>verbal content</i>
	Powaqqatsi	5:59	Feature film; opening scenes from Powaqqatsi, a film about various life struggles <i>no verbal content</i>
	Tea	3:26	Live action skit; surrealistic stop-motion skit about a guy serving his friend a poisoned cup of tea <i>no verbal content</i>
	Tony vs. Paul	5:02	Live action skit; stop-motion animation about two friends who get into a fight and then make up at the end <i>no verbal content</i>

5.6.1. VIDEO SELECTION PROCESS

The Cartoon study videos were selected from Channel Frederator, an online cartoon podcast. Videos from this source were chosen because they were all animated cartoons that we felt would appeal to a wide audience.

The Text vs. Audio study videos were chosen using a more rigorous process because they needed to satisfy criteria important to the study design. The videos could not

be offensive, and they had to either contain verbal content (i.e., people speaking) or not contain any verbal content. In the non-verbal case, videos could only contain instrumental soundtracks. A round of pre-testing was used to collect ratings and evaluations of videos. These ratings and evaluations were used to pick the final set of Text vs. Audio study videos, listed in Table 5-2.

5.6.2. VIDEO RATINGS

In each study, participants were asked to rate the videos they watched on a 5-point scale⁵. This question was asked as “After each cartoon finishes, please rate it below (circle your rating, 5 is the highest)” (Cartoon), “On a scale of 1-5, how entertaining was this video?” (Text vs. Audio pretest), “After each video finishes, please circle your rating below (5 is the highest)” (Text vs. Audio), and “Please rate this video from 1 to 5 stars (5 stars is highest)” (Tag Evaluation, Best Parts). A summary of the mean ratings for each video in each study is given in Table 5-3. Standard deviations are listed in parentheses. For the videos in the Text vs. Audio pretest, the number of people who rated each video is shown in square brackets. For all other studies, the number of participants is shown in the table header.

Table 5-3. Ratings for each video across the empirical studies. All ratings are on a 5-point scale, with 5 as the highest rating. Standard deviations are listed in parentheses. Missing values indicate that a video was not used in the corresponding study. For the Text vs. Audio pretest, the number of raters for each video is listed in square brackets.

	Cartoon	Text vs. Audio (pretest)	Text vs. Audio	Tag Evaluation	Best Parts
	N = 85	[N]	N = 144	N = 30	N = 20
Cartoon study videos					
Emerge	2.4 (1.2)				2.0 (1.0)
Fuggy Fuggy	3.1 (1.3)			2.9 (.87)	2.5 (1.1)
In the Rough	3.6 (1.1)			3.7 (1.0)	3.4 (1.2)
Pen Pals	4.0 (1.1)			3.8 (1.0)	4.1 (.85)
Penguin’s Christmas	3.8 (.97)				3.6 (.99)
Plumber	3.4 (1.3)			3.7 (.99)	3.7 (.97)

⁵ Due to an oversight, participants in the Chat Distraction study were not asked to rate the videos they watched.

	Cartoon	Text vs. Audio (pretest)	Text vs. Audio	Tag Evaluation	Best Parts
	N = 85	[N]	N = 144	N = 30	N = 20
War Photographer	2.7 (1.5)				2.4 (1.2)
Text vs. Audio study videos					
Ali G - War		3.6 (.89) [5]	3.3 (1.3)	3.9 (.80)	3.4 (1.2)
Brothas From the Same Motha		1.7 (.71) [9]	2.1 (1.1)	2.2 (1.1)	2.1 (1.1)
Daughters		3.7 (1.4) [6]	3.5 (1.3)	3.6 (1.1)	3.1 (1.6)
Gopher Broke		4.6 (.79) [7]	4.2 (.98)	3.7 (1.2)	3.7 (1.3)
Paddy the Pelican		1.7 (.82) [6]	1.9 (.99)	1.8 (.97)	1.6 (.81)
Powaqqatsi		1.0 (0.0) [3]	2.4 (1.2)		2.0 (.97)
Tea		2.1 (.88) [10]	2.3 (1.1)		1.6 (.94)
Tony vs. Paul		4.1 (1.1) [7]	3.6 (1.2)	3.3 (1.3)	2.9 (1.4)

Overall, the videos tended to receive stable ratings across studies. Eight videos had a range between the highest and lowest mean ratings of greater than half a point on the rating scale. These videos were “Fuggy Fuggy” (2.5 - 3.1), “Ali G - War” (3.3 - 3.9), “Brothas From the Same Motha” (1.7 - 2.2), “Daughters” (3.1 - 3.7), “Gopher Broke” (3.7 - 4.6), “Powaqqatsi” (1.0 - 2.4), “Tea” (1.6 - 2.3), and “Tony vs. Paul” (2.9 - 4.1). This observation merely reinforces the point that peoples’ preferences for content is widely diverse, and our attempts to control the quality of content across the studies generally only worked at a coarse level.

5.7. SUMMARY AND CONCLUSIONS

- This dissertation uses a combination of field and laboratory studies to improve our understanding of collaborative online video watching.
- Field studies provide opportunities to observe real-world behaviors at the cost of losing the ability to accurately measure aspects of the experience (e.g., distraction).
- The controlled environment of the laboratory enables causal inferences to be made at the cost of assuming that the behaviors observed in the lab are reflective of those that would normally occur in the real world.

- Measures important to the studies in Part II are distraction, enjoyment, and sociability. The scales used to measure these constructs are listed in Appendix B.
- The videos used in the studies were reflective of several popular types of short video clips found on sites like YouTube. These videos give the studies a degree of ecological validity because they are reflective of a real-world online video experience.
- Some videos were enjoyed more than others, providing us with the opportunity to observe how watching collaboratively affects enjoyment of content.

6.

THE MOVIELENS STUDY⁶

We begin our exploration of collaborative online video watching by asking a fundamental question: will people enjoy chatting while watching a movie? If chatting while watching is not enjoyable, then the sociability argument has no credence – social interaction cannot augment television simply because people won't interact while watching. Hence, it is important to understand whether chatting while watching is an activity in which viewers will partake. The best way to answer this question is in the context of a real-world study, in which viewers are less inclined to do what the experimenter tells them. This way, we can be more confident that participants' behaviors (i.e., usage of the chat feature) reflect how they would act in other similar situations (i.e., on other online video sites).

MovieLens is an online community where people rate and discuss movies (Miller et al. 2003). To gain insight into the question of whether or not people would chat while watching a video, we hosted a series of “movie nights” for MovieLens users. At these movie nights, users watched full-length feature films and chatted live with one another using the End System Multicast software (Chu et al. 2001). These movie nights were hosted during February and March of 2006⁷.

⁶ The results reported in this chapter appear in part in (Weisz et al., 2007).

⁷ For historical perspective, this study was run exactly one year after the founding of YouTube, about eight months before the founding of Justin.TV, and about four or five months before the founding of UStream.TV. At that time, the idea of combining live chat with video was nascent.

6.1. RESEARCH QUESTIONS

This study addresses two primary research questions: will people chat while watching a video (and what will they chat about), and will they find the chat distracting?

RQ 6-1: Will people chat while watching a video? What will they talk about?

RQ 6-2: Is it distracting to chat while watching?

6.2. DESIGN

Two experimental conditions were used in this study: a chat condition to understand how people experience chatting while watching, and a no-chat condition as a control for making comparisons.

6.3. PARTICIPANTS

We recruited 65 MovieLens members to participate in this study through a series of email invitations and advertising on the MovieLens web site. Of these participants, 24 completed the study by tuning into at least one movie showing, resulting in a dropout rate of 37%.

Participants were randomly assigned to have the chat feature while watching. Fifteen participants with the chat feature completed the study (62.5%).

6.4. METHOD

We showed five feature films to participants over the course of four weeks. Each film was two to three hours long. Because the movie showings required tuning into the broadcast at a specified time, each movie was shown several times – ranging from two to six times – to accommodate participants in different time zones. On average, each participant joined 2.1 (SD = 1.1) movie showings.

Participants watched the films using the End System Multicast (ESM) software (Chu et al., 2001). ESM is a peer-to-peer video broadcasting system that enables hundreds of viewers to simultaneously tune into a streaming video broadcast. For this study,

ESM was enhanced with a user interface that worked on Mac OS X and Windows. Targeting these operating systems allowed us to reach the widest possible audience. ESM was also equipped with a text chat feature that enabled participants to chat with each other. Figure 6-1 shows a screenshot of the ESM software and its chat feature⁸.

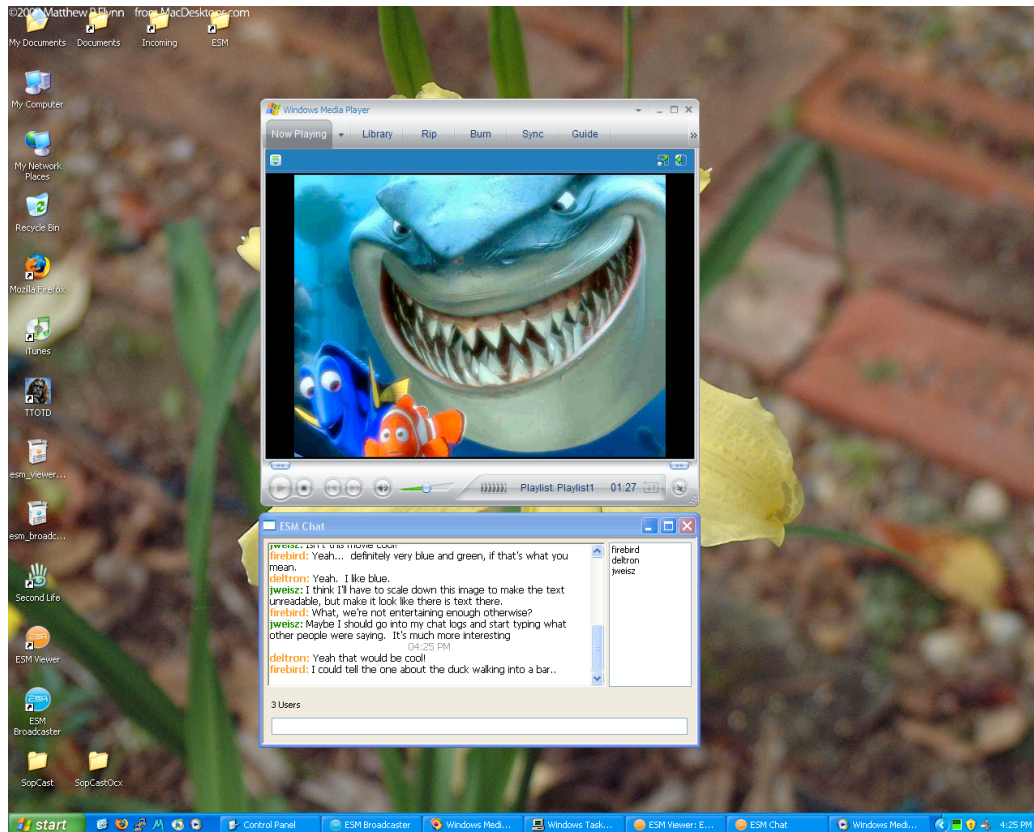


Figure 6-1. Screenshot of the ESM software (minimized) playing a movie in Windows Media Player. The text chat feature (shown) enabled simultaneous viewers to chat with each other while watching.

Participants with the chat feature chatted with each other in small groups. During each of these movie showings, I was also present in the chat to help participants debug the ESM software and talk to participants when only one showed up for the showing. By participating in the experiments, I was able to make sure that participants had someone to chat with. I followed a script for interacting with participants that included greeting them when they joined the movie showing, responding to their inquiries, and prompting them to chat when they were quiet for

⁸ Due to an oversight, no screenshots were taken during the MovieLens study. The screenshot in Figure 6-1 is of a testing session.

more than 10 minutes. I also provided technical support for ESM when participants had difficulties tuning into the video or experienced glitches during the video.

6.5. MEASURES

Participants filled out a survey after each movie showing. This survey asked participants to describe their use of the chat feature, rate their enjoyment of the movie and the chat, and rate whether they felt distracted from chat.

6.6. RESULTS

6.6.1. CHAT AND DISTRACTION

Overall, participants chatted in this study. Participants in each movie session produced an average of 190 lines of chat over the course of a two to three hour movie. This amount of chat corresponds to about 1.1 lines of chat each minute. Thus, although chat occurred at a relaxed pace, participants nonetheless chatted.

After each video, participants with the chat feature were asked to rate their enjoyment of the chat and whether the chat distracted their viewing. These ratings were made on a 5-point Likert scale that asked the degree to which participants agreed with the statements “The chat made this movie more enjoyable to watch” and “I was distracted by the chat.” Overall, participants tended to agree that chat made the movie more enjoyable to watch ($M [SD] = 3.3 [.68]$), and tended to disagree that they were distracted by chat ($M [SD] = 2.6 [.89]$).

Participants also rated the movies on an 11-point star scale, with points at each half-star increment (i.e., .5 stars, 1 star, 1.5 stars, etc.). This scale corresponded to the movie rating scale on MovieLens. Overall, the movies were rated 4.0 stars ($SD = .90$ stars). Controlling for participants who watched multiple movies, participants with the chat feature did not rate the movies significantly differently than participants without the chat feature (chat: $M [SD] = 3.9 [.96]$ stars, no chat: $M [SD] = 4.2 [.77]$ stars, $F [1,21.35] = .81$, $p = n.s.$).

To qualitatively understand the experience of chatting while watching, participants were asked to comment on their experiences. Some participants greatly enjoyed the chat, and felt that it helped their understanding of the video content.

"It was very fun - it was helpful that someone who actually understood the movie could help me understand it - very much increased my enjoyment of it." (ML1)

"I'm also responding positively to the notion of there being a community of people out there sharing my experience." (ML3)

"For me the chat feature was a big part of what made me tune in to the movies ... If the chat hadn't been there, I think I could just as well watch a movie on the TV, or downloaded a movie in advance." (ML14)

Other participants felt that the chat was distracting, and one participant (ML2) did not see any value to the chat feature at all.

"I'm not interested in chatting online, especially not during watching a movie." (ML6)

"[I] disliked that [the chat] was somewhat distracting; had there been more chatter it could have become annoying." (ML3)

"[I] just didn't find it possible to concentrate on movie and chat. If I'm watching a movie, I don't want/need other stimuli." (ML2)

Finally, one participant felt that, with practice, the distraction might become less bothersome.

"[I] don't find it too distracting--I'm taking an online class where we have audio and chat going at the same time, so I'm getting used to multitasking like this" (ML13)

Other distractors existed in this study as well, as a consequence of this being a real-world study. Participants reported doing other activities while watching, including email or instant messaging (67%), browsing the web (62%), and talking on the phone (29%). Interestingly, more participants with chat did email or instant messaging than participants without chat (chat: 86%, no chat: 14%, $\chi^2(1, N=53) =$

14.1, $p < .001$), and more participants with chat browsed the web than participants without chat (chat: 81%, no chat: 18%, $\chi^2 (1, N=52) = 8.1, p = .004$).

6.6.2. CHAT TOPICS

Participants chatted about a variety of topics, including the movie they were watching, the study in which they were participating, and the ESM technology used to run the broadcast. Chat transcripts were coded to understand the distribution of topics during the movie showings. The unit of analysis for the coding was at the block level. Each block consisted of a sequential series of chat messages that were in close temporal proximity to each other and part of the same conversational thread. Only one coder (the author) performed the blocking and coding; thus, the results are not meant to be statistically meaningful, just qualitatively descriptive⁹. The distribution of chat topics is shown in Table 6-1.

Table 6-1. Distribution of chat topics across the movie showings.

Category	Percentage of chat
Personal topics	32
Movie-related chat	28
The MovieLens study	23
The ESM technology	11
Technical support & troubleshooting	6

6.7. DISCUSSION

The MovieLens study was designed to answer two preliminary questions about chatting while watching: would people do it (RQ 6-1), and would they find it distracting (RQ 6-2). The answer to both of these questions is “yes.” Participants did chat with each other in the study, and to some degree found it distracting. Yet, the extent to which participants were distracted and whether this distraction had a negative impact on their experience is unclear. Participants reported enjoying the movies they watched, and they did not seem to feel overly distracted by the chat. The fact that participants self-distracted by engaging in other activities like instant

⁹ The real value of the coding scheme developed in this study was that it served as a basis for the coding scheme discussed in Chapter 7.

messaging or talking on the phone supports the argument for including chat with video; curiously, both of these activities are also socially-engaging. However, self-distraction was done more by participants with chat. It is possible that participants with chat were primed to engage in other activities, simply by having the chat feature.

As for the topics of chat, there was a distinction between off-topic chat (e.g., personal topics) and on-topic chat (e.g., movie-related chat). Although the proportion of chat in each category was biased because the experimenter participated in the conversations, we nonetheless see that participants were willing to chat about themselves and their personal lives while watching the movies. In addition, the movies did provide content to discuss in the chats. These findings encourage the inclusion of a chat feature with online video for the purpose of building community.

This study shows that participants generally enjoyed chatting while watching the videos, although some participants reported being distracted from the chat. The Chat Distraction study, discussed in Chapter 7, focuses on quantitatively measuring the extent to which participants are distracted by chat while watching videos.

6.8. SUMMARY AND CONCLUSIONS

- MovieLens users were recruited to watch movies and chat with each other during a series of movie nights, hosted from February to March, 2007.
- Twenty-four participants tuned into the movies. Fifteen were assigned to have a text chat feature available to use.
- Participants with the chat feature generally used and enjoyed it. Some participants reported that it was distracting to chat while watching.
- Chat during the movies included topics related to the movies as well as personal topics. This finding led to the development of a more detailed examination of chat topics in the Cartoon and Text vs. Audio studies (Chapters 8 and 9).
- The findings in this study provide evidence supporting the sociability argument, as participants had fun chatting with others while watching, and they chatted about personal topics.

- The findings in this study also provide evidence supporting the distraction argument, as some participants reported being frustrated from multitasking between watching the movie and chatting.

7.

THE CHAT DISTRACTION STUDY¹⁰

The human factors literature discussed in Chapter 1 suggests that chatting while watching videos will be distracting for viewers because of the need to attend to multiple simultaneous sources of information (Wickens & Hollands, 2000). These sources can be visual, in the case of text chat, or auditory, in the case of voice chat.

In Chapter 6, we saw that some participants in the MovieLens study reported feeling distracted from using a text chat feature while watching movies together. In this chapter, I quantify the degree to which viewers are distracted by attending to a text chat. I also examine a simple tweak to the user interface that may (but ultimately does not) reduce distraction.

7.1. SIMPLIFYING TEXT CHAT

One reason why chat is distracting while watching a video may have to do with the presentation of chat on-screen. Many online video sites that incorporate text chat use an IRC-style display metaphor in which chat messages are rendered in a text box and new messages are appended at the bottom (Figure 7-1). These chat boxes display some amount of chat history, depending on both the size of the chat box and the font used to display the chat. This chat history may be distracting, because it may present more information than users need to conduct a conversation. In addition, updates to the interface may not be salient enough. For example, to detect when a new message arrives, users must either register the scrolling motion of the chat in their peripheral vision (if attending to the video), or poll the chat box by periodically

¹⁰ Portions of this work have previously appeared in (Weisz, 2009).

looking between it and the video. When viewers do detect motion in their periphery, they must saccade and focus their vision on the chat box to the point at which the new message is located. If several new messages have been added, the user may need to scan the list to find the new content, as well as refresh their memory of the conversation. This process of detecting, finding, and processing new content takes time away from watching the video, and may contribute to the overall level of distraction felt from chat.

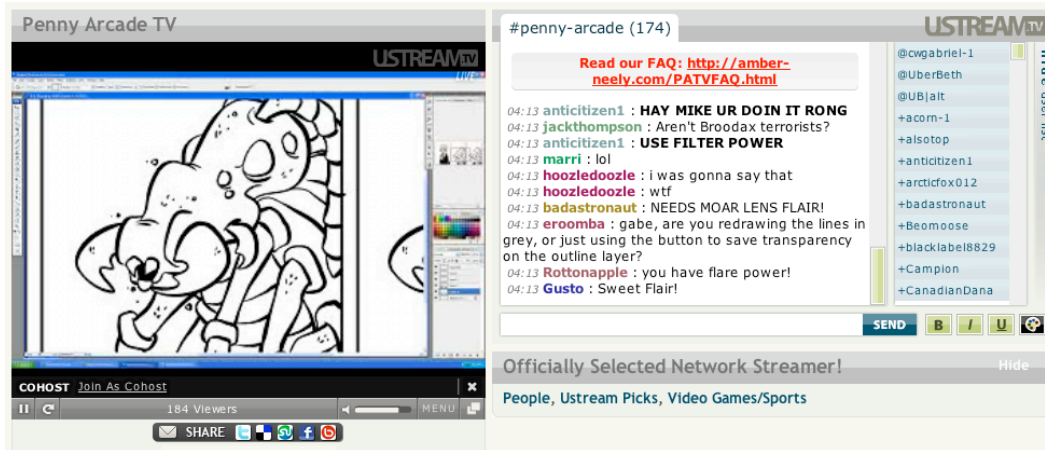


Figure 7-1. Screenshot from UStream.TV showing video and text chat.

An alternative method for presenting chat is to remove the chat history altogether. Removing the chat history ensures that new chat is always placed in the same location, eliminating the need to actively search for the location of new content (although it does not eliminate the need to poll for its arrival), and removes the ability for viewers to spend time searching for new content. However, removing chat history may have potential negative social consequences. If viewers fail to recognize the appearance of a new chat messages, their ability to carry on an extended back-and-forth conversation may be hampered. Therefore, there is a potential trade-off in this scheme; although it may reduce distraction by reducing the amount of visual scanning needed to find new chat messages, it may also reduce the quality of social interactions if viewers miss pieces of the conversation. Further, an increase in chat message traffic from, for example, the conversational repairs that occur because of missed chat messages, may also increase both distraction and frustration.

The removal of chat history is somewhat motivated by the design of Coterie, an alternative interface for IRC-based chat rooms (Spiegel, 2001). In Coterie, chatters are represented by ovals at the bottom of the screen, and their chat messages

emanate from their oval and move upwards over time, until they disappear off-screen. Coterie lacks a historical summary of chat; once a chat message disappears, it cannot be retrieved for review. Spiegel argues that this behavior is in fact a ‘problem’ for the Coterie visualization. In this chapter, we evaluate whether the lack of a chat history really is a problem by measuring whether video viewers are able to recall a conversation without history while watching a video.

7.2. RESEARCH QUESTIONS

The research questions addressed in this chapter revolve around the distracting effects of chat on the video-watching experience.

RQ 7-1: How distracted are viewers who read chat messages while watching a video?

RQ 7-2: Does removing the chat history log help reduce distraction? Does it impair viewers’ ability to recall the chat?

We answer these questions in the context of a simulated laboratory study. In this study, participants watch videos and read chat, but do not chat themselves. The chat they read is from a pre-recorded chat transcript created specially for this study. Thus, all participants read the same chat messages while watching the videos. This level of control allows us to objectively measure the distractive effects that chat has on the video, as well as the distractive effects that the video has on chat.

7.3. DESIGN

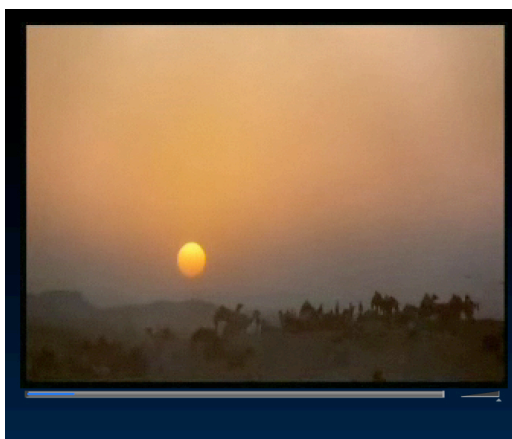
This study compares three chat interfaces: one with a full chat history log (full chat history), one that only displays the last line of chat (no chat history), and a control condition without any chat (no chat). Figure 7-2 shows screenshots of the three conditions.



(a)



(b)



(c)

Figure 7-2. Screenshots from the Chat Distraction study. (a) Full chat history condition. (b) No chat history condition. (c) No chat condition.

7.4. PARTICIPANTS

Forty-two participants were recruited for this study from the CBDR web site¹¹. The average age of the participants was 25.2 years (SD = 9.5 years). Eighteen participants were female and 24 were male. Participants were paid \$15 for their participation, which took approximately one hour.

7.5. METHOD

Participants were shown a sequence of eight videos in random order. These videos were the Text vs. Audio study videos, described in Section 5.6.

This study was designed with two goals in mind. The first was to quantitatively measure the extent to which people are distracted while watching a video and chatting. The second was to test whether or not eliminating the chat history would have an effect on levels of distraction. In order to control the amount of chat seen by each participant, we created a pre-recorded chat transcript for each video. Thus, our quantitative measure of distraction – memory – is reflective only of reading chat, and not producing it.

The presentation of chat messages was synchronized for each video. Thus, chat that occurred during “Gopher Broke” was always shown during “Gopher Broke,” even if one viewer watched this video first and another watched it last. Synchronization of chat to the video ensured that any references made to the video in the chat were correctly aligned to the video.

7.6. MEASURES

The primary measure in this study was memory, which objectively quantified how much attention participants paid to the video and the chat. To measure distraction of the chat on the video, we asked two questions about the content of each video. To measure distraction of the video on the chat, we asked two questions about the content of each chat. Thus, four questions were asked about each video. Examples of video and chat memory questions are given in Appendix B. To avoid difficulties in recall, these questions were asked immediately following each video.

¹¹ Center for Behavioral and Decision Research. <http://www.cbdr.cmu.edu/>

The number of correct questions about the videos was summed and labeled the video memory score. Similarly, the number of correct questions about the chat was summed and labeled the chat memory score. The maximum attainable values for each of these scores was 16. Participants in the no-chat condition did not have a chat memory score.

7.7. RESULTS

The no chat condition was a control condition to verify that no ceiling effects were present in the video memory score. The mean video memory score for participants without chat was 13.3 (SD = 1.39) of 16 questions correct. No participants had a perfect video memory score. Therefore, we do not find evidence for a ceiling effect in video memory. Further, no participants in the chat groups had a perfect chat memory score, so we also find no evidence for a ceiling effect in chat memory.

Participants with full chat history had a mean video memory score of 11.1 (SD = 2.0) questions correct (69.4%). Participants with no chat history had a mean video memory score of 11.3 (SD = 2.4) questions correct (70.6%). The difference between these two conditions was not significant ($F [1,39] = .08, p = \text{n.s.}$).

Comparing the video memory scores of participants with and without chat, we found that participants with chat recalled about two fewer questions about the videos than participants without chat (chat: $M [SD] = 11.2 [2.15]$ questions correct, no chat: $M [SD] = 13.3 [1.39]$ questions correct). This difference was significant, $F (1,39) = 10.7, p < .002$.

Participants with full chat history had a mean chat memory score of 10.3 (SD = 3.0) questions correct (64.3%). Participants with no chat history had a mean chat memory score of 9.1 (SD = 2.3) questions correct (56.9%). This difference was not significant ($F [1,26] = 1.3, p = .27$). Therefore, the elimination of the chat history did not affect recall of the chat.

Finally, there was a weak but insignificant correlation between the video memory scores and the chat memory scores ($r = .27, p = .17$). Thus, participants who remembered more about the videos also tended to remember more about the chats.

7.8. DISCUSSION

The participants in this study with chat remembered less about the videos than the participants without chat – enough that they incorrectly answered two more questions than participants without chat. Thus, this study objectively shows the distractive effects of text chat on watching a video (RQ 7-1). As this experiment only measured distraction from reading others' chat, we speculate that the amount of distraction would increase when people are invested in the chat, and they are expected to compose chat responses to each other.

The hypothesis that eliminating chat history would reduce the distractive effects of chat on video was not borne out by the data (RQ 7-2). Participants with no chat history log recalled the same amount about the videos as participants with full chat history. Interestingly, participants in these conditions also recalled the same amount about the chats – about 55-65% – suggesting that eliminating the chat history log may not be detrimental to a viewer's ability to recall the conversation.

Recall of the videos was higher than recall of the chats, suggesting that participants did not pay as much attention to the chat as they did to the videos. Even when participants had the full chat history log, and could review what was said at their own pace, it seems that they did not do this because their chat memory scores were not higher than participants without the history log. One explanation is that participants were not motivated to remember as much as they could from the chats simply because were not actively participating in them and thus were not invested in them. However, participants were instructed that their memory of the chats (and the videos) would be tested. Thus, we conclude that the videos required more attentional resources to process than the chats, and the videos distracted participants from the chats even when the full chat history was available for review.

Interestingly, the correlation between the video memory and chat memory scores was positive. With a negative correlation, we might conclude that attention to the video shifted attention away from the chat, and thus participants would remember more from the videos and less from the chat (or visa versa). However, with a positive correlation, it seems that the more attention each participant put into the entire experience – watching both the chat and the video – the more they remembered from it.

In the next chapter, we examine the distractive effects of chat on the video in a social context, in which viewers engage in the act of chatting with each other. Because our manipulation of the user interface did not seem to reduce distraction from chat, we examine another method for reducing distraction. This method restructures the viewing experience to give viewers an opportunity to chat without being distracted by the video. This restructuring comes in the form of intermissions added in between a series of videos.

7.9. SUMMARY AND CONCLUSIONS

- This chapter reports on an experimental laboratory study in which 42 participants watched a series of videos while reading a pre-recorded transcript of chat.
- Three user interfaces were compared: one that contained a full chat history log, one that only showed the last line of chat, and one that did not show any chat. It was expected that removing the full chat history log would reduce distraction by making it easier for participants to find new chat when it arrived.
- Distraction was measured by asking participants to recall specific things from the videos and from the chats.
- Participants without chat recalled more than participants with chat. There was no difference in recall between participants with the full history log or participants with only the last line of chat.
- This study finds objective evidence that simply reading chat while watching videos is distracting.

8.

THE CARTOON STUDY¹²

The Chat Distraction study (Chapter 7) shows that viewers are distracted when they read chat while watching a video. This result supports the distraction argument, that chatting while watching a video will not be enjoyable because it is too distracting. In the MovieLens study (Chapter 6), some participants reported not enjoying chat for precisely this reason. The first goals of this chapter are to further explore the distractive effects of chat on the video, and to evaluate another strategy for reducing distraction.

This new strategy for reducing distraction is to restructure the viewing experience so that viewers do not need to simultaneously attend to video and chat. This way, only one visual (or auditory) source of information needs to be processed at a time. To add this structure, we can simply create a break period during which viewers can chat with each other after a video has finished playing. Adding break periods can be useful for collaborative online video sites in a variety of ways: advertisers may want to show commercials during the breaks, community leaders may want to hold discussions, or viewers may simply want to chat amongst themselves. Break periods are also a realistic method for reducing distraction, as television content is already primed with periods for commercial breaks. In fact, much of the major-network television content online already has commercial breaks for advertising (e.g., TV shows on Hulu).

Another question revolving around the issue of distraction is how it trades-off with the social experience of watching with others. Does the value viewers derive from

¹² The results in this chapter appear in part in (Weisz et al., 2007) and (Weisz, 2009).

the social experience of watching and chatting with other people outweigh the distractive effects of having the chat feature? If so, then the issue of distraction may be less important; the human factors literature explains why viewers are distracted, but if the distraction is not ruinous to the experience, then we need not worry about it when designing a collaborative online video experience. Thus, the second goals of this chapter are to understand the social value viewers derive from chatting while watching and understand whether it is greater than the frustration experienced from being distracted.

The social value derived from chat is important to study because it gives insight to the merits of the sociability argument – that chatting while watching can help build social capital because viewers are not watching alone. There are several ways in which to examine social value. The first is the impact that chatting while watching has on viewers' feelings toward one another. If viewers do not feel the social presence of other viewers, then collaborative watching may not be able to provide a sociable experience.

Another way to examine social value is by examining the topics of chat. In online communities, bonds between community members typically form as members chat with each other about off-topic subjects, such as their personal lives (Sassenberg, 2002; Preece & Maloney-Krichmar, 2003; Ren, Kraut, & Kiesler, 2007). If viewers remain on-topic in their chats, by only talking about the video they are watching, then the likelihood of building social relationships is diminished. This situation may occur if the presence of the video is so strong that it either discourages people from talking, or discourages them from talking about anything else.

Finally, there is a subtlety to the sociability argument that must be addressed. Social capital must both be built, by creating new relationships and ties to other people, and it must also be maintained, by periodically interacting with existing members in one's social network (Putnam, 1995; Putnam, 2001). Thus, we must compare between two types of viewers in our study: groups of friends who already know and have established relationships with one another, and groups of strangers who have not met before. These two groups of viewers mirrors the social dynamics present in many online communities. Some community members know each other already, and some are newcomers meeting others for the first time. Studying groups of friends and strangers separately lets us understand whether collaborative online video watching can be used to maintain existing relationships as well as create new ones.

8.1. RESEARCH QUESTIONS

The overarching research questions in this study are about distraction and sociability.

First, we ask whether the social viewing experience is enjoyable, and if it is more enjoyable than the solitary viewing experience.

RQ 8-1: Is it fun to chat while watching? Is it more fun than watching alone?

Next, we ask the extent to which viewers feel distracted while watching (i.e., subjective feelings of distraction). In addition, we ask whether break periods will reduce viewers' feelings of distraction, and if viewers take advantage of them by chatting more during the break periods than the videos.

RQ 8-2: How distracted do viewers feel when chatting while watching videos? Do break periods help reduce their feelings of distraction? Do they restructure their interactions to take advantage of the break periods?

To answer questions about the social value of chat, we must examine what viewers talk about while watching and whether there is a quantifiable impact of watching together on viewers' feelings toward one another. Note that we are not asking directly whether chatting while watching increases social capital; that question is not answerable in the context of a laboratory study. Instead, we ask whether certain preconditions of relationship formation – liking other people and feeling their presence – are present in the momentary experience of collaborative online video watching.

RQ 8-3: What do people talk about while watching a video together?

RQ 8-4: Does chatting while watching have an impact on viewers' feelings toward one another?

These questions are answered in the context of a quasi-experimental laboratory study.

8.2. DESIGN

In this study, groups of friends and groups of strangers watched videos and chatted with each other in the laboratory. To examine methods for reducing distraction from chat, we compared between two types of break periods: intermissions between a sequence of videos, and a break at the end of the entire sequence of videos. In addition, groups without chat are used as a control. Thus, this study uses a 3 x 2 factorial design with break type (no chat, chat with intermissions, chat with an end break) and group composition (friends or strangers) as between-subjects factors. This study was quasi-experimental in nature due to the fact that group composition required pre-existing groups and could not be randomly assigned. Break type was assigned randomly.

8.3. PARTICIPANTS

Participants were recruited in groups of two to four people from the CBDR web site. Groups averaged 2.8 members. Thirty groups were recruited, with a total of 85 participants. Groups were distributed evenly across the six combinations of experimental conditions, with five groups per combination of group composition and break type.

To recruit groups of friends, we asked that people interested in the experiment find two friends to participate with them. To recruit strangers, participants simply signed up for one of the time slots we offered. Group size did not differ significantly among the experimental conditions ($F [3,16] = .6, p = n.s.$).

The average age of the participants was 24.3 ($SD = 7.3$) years, and approximately half were female. Seventy-five percent of participants were students, five percent were faculty or staff, and the rest were alumni, retired, or did not list their affiliation. Participants were paid \$15 each for their participation, which took approximately one hour.

8.4. METHOD

Participants in each group watched a series of cartoons on the computer. Participants in the intermission and end break conditions could communicate with each other using a text chat feature.

Due to the nature of the laboratory in which this experiment was conducted, participants were seated in the same room. Participants sat in cubicles that provided a visual separation from one another. Although participants remained in auditory range to one another, they wore headphones to hear audio from the computer. No group chatted aloud to one another, although several participants did laugh out loud during the study.

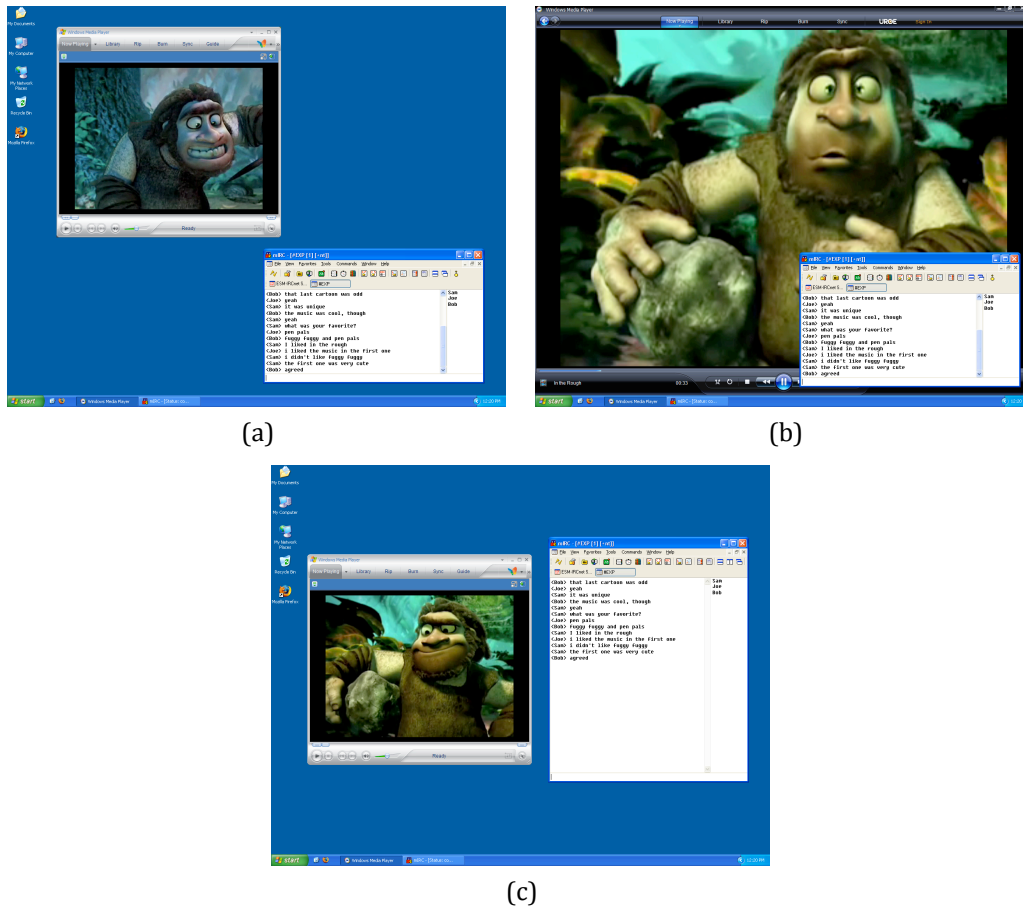


Figure 8-1. Different arrangements of video and chat windows. Participants were allowed to move and resize the windows to their individual preference. (a) Default arrangement with chat offset from video. (b) Full-screen video with chat overlaid. (c) Chat and video side-by-side.

As this study was conducted using laboratory machines on a LAN, we used Windows Media Player to play streaming video from a server on the LAN. The streaming video ensured that everyone watched the same content at the same time. We used the mIRC IRC client for chat. Chat logs were collected using an open-source IRC server instrumented to log timestamps, message senders, and the contents of each line of chat. Participants were allowed to rearrange the positions and sizes of the chat and

video windows to their liking, and about one-third of participants did this. Figure 8-1 shows screenshots of several popular arrangements.

Participants watched the seven Cartoon study videos discussed in Section 5.6. Each cartoon lasted between three and six minutes. In the intermission condition, one-minute intermissions were placed between each cartoon. In the end-break condition, participants were given a six-minute period for discussion at the end of the cartoons, keeping the break time equal with the intermission condition. None of the participants had previously seen any of the cartoons before.

In both chat conditions, participants were told that they could chat with the other participants at any time during the study (cartoons or breaks), about any topic. Participants were only told of the availability of the chat feature, not that it was a mandatory requirement of their participation. In the no-chat condition, participants did not have the chat feature and thus did not receive the breaks.

8.5. MEASURES

All participants rated each cartoon immediately after it had finished, to avoid difficulties in recall. These ratings were made on a 5-point scale, representing how much they liked each cartoon. All participants also completed a final survey that asked about their experience, the chat feature and feelings of distraction, and the other people in their group; questions specific to the chat feature were omitted for participants without chat. These questions were used to build four scales. They were enjoyment (2 items, Cronbach's $\alpha = .93$), chat enjoyment (3 items, $\alpha = .89$), liking (4 items, $\alpha = .81$), and closeness ($N-1$ items for N group members, $\alpha = .84$). Details of these scales are given in Appendix B.

Two measures of sociability were used in this study. The liking scale asked participants questions about how much they liked their other group members. The closeness scale, based on the Inclusion of Other in Self scale (Aron et al., 1991), asked participants to rate how close they felt to one another. Participants were asked to rate their feelings of closeness to each other member of their group; for analysis, these values were averaged.

8.6. RESULTS

Unless otherwise specified, the primary method of analysis for the outcome variables in this study was an analysis of variance (ANOVA). In the ANOVA models, the explanatory variables were break type (no chat, chat with intermissions, chat with an end break) and group composition (friends or strangers). Group ID was also included in the model as a random effect to control for within-group variance; this random effect also makes a correction to the within-groups (error) degrees of freedom term, resulting in real-valued degrees of freedom for some cases. Contrast testing was used to compare between specific conditions (e.g., intermissions vs. end break), as well as between groups that had the chat feature and groups that did not.

8.6.1. FUN, ENJOYMENT, AND VIDEO RATINGS

Enjoyment. The first research question asks if it is fun to chat while watching, and if it is more fun than watching alone (RQ 8-1). Overall, participants enjoyed watching the cartoons, as the mean enjoyment score was 4.0 (SD = .84) out of 5. Enjoyment was significantly correlated with participants' average cartoon ratings ($r = .47, p < .001$). Enjoyment did not differ between the chat conditions ($F [1,79] = 1.01, p = \text{n.s.}$); thus, participants without chat enjoyed the cartoons as much as participants with chat (no chat: $M [SD] = 3.8 [.98]$, chat: $M [SD] = 4.0 [.76]$). There was no correlation between the amount an individual chatted and their enjoyment of watching the cartoons ($r = .14, p = \text{n.s.}$).

Chat enjoyment. Participants with the chat feature enjoyed using it while watching the cartoons, as the mean chat enjoyment score was 4.2 (SD = .64) out of 5. Participants with intermissions did not differ from participants with the end break in their enjoyment of chat (intermissions: $M [SD] = 4.2 [.76]$, end break: $M [SD] = 4.2 [.52]$; $F [1,53] = .15, p = \text{n.s.}$). Groups of friends enjoyed the chat feature more than groups of strangers (friends: $M [SD] = 4.4 [.56]$, strangers: $M [SD] = 4.0 [.67]$; $F [1,53] = 5.67, p = .02$). The interaction between group composition and break type on chat enjoyment was not significant ($F [1,53] = .15, p = \text{n.s.}$).

There was a significant correlation between the amount an individual chatted and chat enjoyment ($r = .26, p = .05$). Thus, either people who chatted more enjoyed the chat more, or people who enjoyed the chat more chatted more.

Cartoon ratings. The cartoons received a mean rating of 3.3 out of 5 (5 highest). A principle components analysis of the cartoon ratings reveals three components with an Eigenvalue greater than 1: two “poor” cartoons (M [SD] = 2.5 [1.1]), three “okay” cartoons (M [SD] = 3.3 [1.1]), and two “good” cartoons (M [SD] = 3.8 [.70]). Thus, the cartoons expressed different levels of quality.

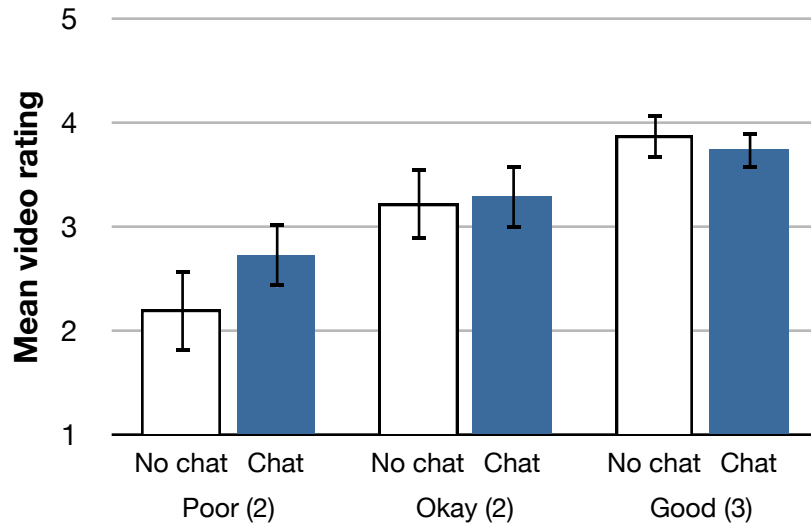


Figure 8-2. Distribution of cartoon ratings for groups with and without chat, by cartoon quality. Error bars represent 95% confidence intervals. The difference in mean rating for poor cartoons was significant ($p = .02$); the differences in mean rating for okay and good cartoons are not significant.

Figure 8-2 shows the distribution of ratings for each level of cartoon quality between groups with and without chat. An ANOVA is used to determine whether groups with or without chat rated each class of cartoons differently. This analysis uses group composition (friends or strangers), break type (no chat, intermissions, end break), and cartoon quality (poor, okay, good) as explanatory variables. We control for within-participant variance by including participant ID as a random effect, and we control for within-group variance by using group ID as a random effect. The model includes all main effects, two-way interactions, and the three-way interaction between group composition, break type, and cartoon quality.

Two significant interactions are seen in this model (the three-way interaction was not significant). First, the ratings of the different-quality cartoons were different for groups of friends and groups of strangers ($F [2,158] = 4.05, p = .02$). Contrast testing reveals that groups of strangers rated the poor cartoons higher than groups of friends (strangers: M [SD] = 2.8 [1.2], friends: M [SD] = 2.3 [1.1]; $F [1,231.3] = 4.75, p$

= .03). No differences were present between friends and strangers for okay cartoons ($F [1,231.3] = 2.21, p = .13$) or good cartoons ($F [1,231.3] = .25, p = n.s.$).

The second significant interaction is between break type and cartoon quality ($F [4,158] = 2.38, p = .05$). This effect can be seen in Figure 8-2, in which the poor cartoons were rated lower by groups without chat than groups with chat. A contrast between the two groups with chat (intermissions and end break) and groups without chat reveals that the difference was significant, $F (1,231.3) = 5.62, p = .02$. Groups with chat rated the poor cartoons higher ($M [SD] = 2.7 [1.2]$) than groups without chat ($M [SD] = 2.2 [1.1]$). Contrasts between groups with chat and groups without chat on okay and good cartoons show no differences (okay: $F [1,231.3] = .04, p = n.s.$; good: $F [1,82.0] = .38, p = n.s.$). Thus, participants who chatted while watching poorer material reported higher levels of enjoyment for that material.

8.6.2. DISTRACTION

Feelings of distraction. The second research question asks how distracted participants felt when chatting while watching. It also asks whether break periods helped them feel less distracted, and whether they took advantage of these periods to chat without being distracted (RQ 8-2). Feelings of distraction were measured on an open 7-point scale, anchored by “not distracted at all” (1) and “very distracted” (7). Participants reported a mean distraction of 3.6 ($SD = 2.0$). Since self-reported feelings of distraction fell in the middle of the scale, we conclude that participants did feel distracted while watching the videos, but not overwhelmingly so. Distraction did not correlate with enjoyment of watching the cartoons ($r = .17$) or chat enjoyment ($r = .12$). However, distraction did correlate with the amount an individual chatted ($r = .38, p = .004$); thus, participants who spoke more in the chat felt more distracted by it.

Break period usage. Participants took advantage of the break periods, conducting roughly 33% of their chat during the breaks, even though the breaks only accounted for about 10% of the time spent in the experiment. However, the majority of chat (62%) occurred during the cartoons, which accounted for about 70% of the time in the experiment. The remaining 5% of chat was typed either before the cartoons began or after they ended.

Breaks and distraction. Participants with intermissions between videos reported feeling less distracted than participants with the break at the end (intermissions: $M [SD] = 3.6 [1.9]$, end break: $M [SD] = 4.1 [2.0]$; $F [1,54] = 3.7, p = .06$). This difference was marginally significant, suggesting that intermissions tended to reduce feelings of distraction from chat. In addition, the mean distraction scores of end-break groups were positively correlated with the amount they chatted during the cartoons ($r = .80, p = .006$), whereas the mean distraction scores of intermission groups were not correlated with the amount they chatted during the cartoons ($r = .07$). Therefore, only the participants in the end-break groups (i.e., without intermission periods) felt more distracted when more chat occurred during the cartoons. Participants with intermission periods tended to feel less distracted overall.

One explanation for the difference in distraction is that groups with intermissions simply chatted less than groups with an end break, and thus felt less distracted. This hypothesis is not supported by the data. Intermission groups did not produce a significantly different amount of chat as end-break groups (intermissions: $M [SD] = 205.1 [161.3]$ lines of chat, end break: $M [SD] = 261.4 [190.6]$ lines of chat, $F [1,18] = .5, p = n.s.$). Further, intermission groups did not significantly differ from end-break groups in the amount they chatted during the cartoons (intermissions: $M [SD] = 129.5 [117.3]$ lines of chat, end break: $M [SD] = 161.3 [134.8]$ lines of chat; $F [1,16] = .43, p = n.s.$). Intermission groups also did not significantly differ from end-break groups in the amount they chatted during the breaks (intermissions: $M [SD] = 66.0 [50.4]$ lines of chat, end break: $M [SD] = 87.5 [70.5]$ lines; $F [1,16] = .79, p = n.s.$).

Break preferences. Introducing intermissions into a sequence of cartoon videos is analogous to introducing commercials in sports programming during breaks in play. Although these commercials take advantage of the natural breaks in the game, they can fragment the experience, and may frustrate viewers who wish the breaks were shorter or nonexistent. We asked participants about their opinions of the break periods, and which type of break they would have preferred. The results are in support of intermissions: 100% of participants with intermissions reported preferring the intermissions, and 52% of participants with the end break reported wanting intermissions. Further, there was no difference in break preferences between groups of friends and groups of strangers ($\chi^2 (1, N=56) = .90, p = n.s.$).

Participants reported wanting flexibility for when they chatted. Of the 57 participants with chat, a majority (63%) reported that they preferred to chat

throughout the entire experience, rather than confining their chat to just the break periods (23%), just the cartoons (9%), or not chatting at all (5%). Thus, participants preferred to manage their own use of the chat feature, rather than have the system impose a rule on when they could or could not chat.

8.6.3. CHAT TOPICS

The MovieLens study showed that viewers chatted about many different topics while watching movies, with a distinction between on-topic chat (e.g., talking about the movie or the study) and off-topic chat (e.g., talking about their personal lives). In this study, we asked participants what their favorite topics of chat were. They included “the cartoons themselves” (participant C16), “the music and the quality of the drawings” (C22), “the rating” (C23), “how good each cartoon was” (C27), and “[the] artistry of videos” (C58).

Participants also made jokes and talked about their lives in their chats.

“We discussed some stuff about our professors by comparing them to the characters. One was related to [two] professors who are a couple and that was hilarious.” (C1)

“I liked chatting with my friends about our inside jokes. It may appear that we don’t like each other, but there is so much love between the three of us that it is hard for a stranger to imagine.” (C4)

Groups of strangers were able to find common ground with each other, and their favorite topics included “information about graduate school” (C13), “smoothies at Lulu’s” (C53, about a local restaurant), and “rating the cartoons” (C56).

To follow up on these informal impressions, we conducted a detailed coding of the chat logs to understand how much participants spoke about the different topics (RQ 8-3). We used the *line of chat* as our unit of coding, although since the content in a single line was not always enough to determine an adequate code, we considered each line of chat in its surrounding context.

We developed our coding scheme iteratively by reading through the chat logs, coding a subset of the chat, and then resolving discrepancies by clarifying the definition of a code, or adding or removing codes. We performed a reliability check with two

independent coders on 12% of the chat corpus and achieved a good inter-rater reliability after four iterations (Cohen's $\kappa = .78$).

The coding categories were: the cartoons, evaluations and ratings of the cartoons, personal topics, laughter, study chat, and greetings and partings. Each line of chat was coded under only one of these categories, except for laughter. As laughter frequently co-occurred with other chat, we assigned multiple codes to these cases. Examples of chat in each coding category, as well as a breakdown of the amount of chat in each coding category, are shown in Table 8-1.

Table 8-1. Distribution of chat categories. Examples of chat are printed in their original form. Lines of chat were coded into only one category, except for laughter. Lines of chat containing laughter were coded as either solely consisting of laughter (7.4%) or containing laughter in addition to other content (9.4%).

Category	Example chat (original form)	% chat
Cartoons	"the colors are pale looks like a bad chinese cartoon of the late 80's" (C21)	41.6
	"Its showing the similarity between weapons and musical ecquipment" (C64)	
	"the dots are supposed to represent human activity and thier choas + beauty" (C59)	
Personal	"im doing sociology and urban studies" (C34)	22.8
	"it is supposed to rain this evening?" (C59)	
	"what's the Catholic deal with seperation...I know divorce is a big no no" (C11)	
Evaluations	"[this] music is awesome" (C20)	13.7
	"hmm so far i actually like the pengiun one the best" (C34)	
	"That was a bit gross although it was a bit funny" (C15)	
Study	"im so happy we're doing this, this is a bonding experiance" (C5)	12.7
	"we only have 2 more [cartoons], Im kinda sad about it" (C45).	
Laughter	":D," "haha," "lol" (and many variations thereof)	7.4 (solo)
	"haha, happy endings are overrated" (C16)	9.4 (mixed)
Greetings & partings	"hi," "hello," "yo," "bye" (and many variations thereof)	1.8

Cartoon, personal, and evaluation chat. A large portion of the lines of chat (41.6%) were about the cartoons, as well as relations of the cartoons to people's lives. For example, participant C59 said, "this is my husband at work on our house,"

and participant C2 said, “we must ask lisa if this is [what] happened to her home today” while watching the cartoon about a plumber.

Participants also spoke about their personal lives during the chats. They talked about their course of study, the weather, sporting events, roommates, and even marriage. About 23% of chat was personal and unrelated to the videos they watched.

We performed ANOVAs at the group level to compare the percentage of chat in each category between the different conditions (group composition and break type). There were no differences among the different conditions in the amount of chat about the cartoons, $F(3,16) = .41, p = n.s.$

For personal chat, the overall differences between groups were not significant ($F[3,16] = 1.99, p = n.s.$). However, end break groups tended to have more personal chat than intermission groups (intermission: $M [SD] = 11.8\% [10.8\%]$, end break: $M [SD] = 24.6\% [17.0\%]$; $F[1,16] = 3.97, p = .06$).

About 14% of the chats were about participants’ evaluations of the cartoons. Strangers chatted about their evaluations of the cartoons about twice as much as friends (30% vs. 15%), but this difference was not statistically significant ($F[1,16] = 2.6, p = n.s.$).

Laughter. Spontaneous laughter occurred frequently. Examples of laughter in the text chat included “lol,” “haha,” and many variations thereof. We also coded happy smilies such as :) and :D as laughter since they were often used to express positive emotions. In total, 7.4% of the lines of chat solely consisted of laughter, and 9.4% of the lines of chat contained some form of laughter.

There was a significant interaction between group composition and break type on the amount of laughter, $F(1,16) = 7.6, p = .01$. Contrast testing shows that groups of friends tended to laugh more when watching with intermissions (friends/intermissions: $M [SD] = 31.6\% [21.6\%]$, friends/end break: $M [SD] = 16.8\% [12.5\%]$; $F[1,16] = 3.34, p = .09$). Groups of strangers laughed more when watching with the break at the end (strangers/end break: $M [SD] = 17.4\% [18.0\%]$, strangers/intermissions: $M [SD] = 3.0\% [4.6\%]$; $F[1,16] = 4.29, p = .05$).

There was no difference in the amount of laughter between groups of friends and groups of strangers, $F(1,16) = .83, p = n.s.$

8.6.4. SOCIABILITY

To compare participants' feelings of liking and closeness to one another, we performed an ANOVA on the liking and closeness measures (RQ 8-4). Explanatory variables were break type (no chat, intermissions, end break), group composition (friends or strangers), and group ID as a random effect to control for within-group variance. Contrast testing between groups with chat and without chat was also performed.

Figure 8-3 shows the mean values of liking between groups of friends and groups of strangers, and groups with and without chat. Liking was measured on a 5-point scale, with 5 being the highest level. Overall, there was a significant main effect of chat on liking, $F(1,78) = 21.8, p < .001$. Participants with the chat feature liked their other group members more than participants without the chat feature (chat: $M [SD] = 4.2 [.73]$, no chat: $M [SD] = 3.5 [.70]$).

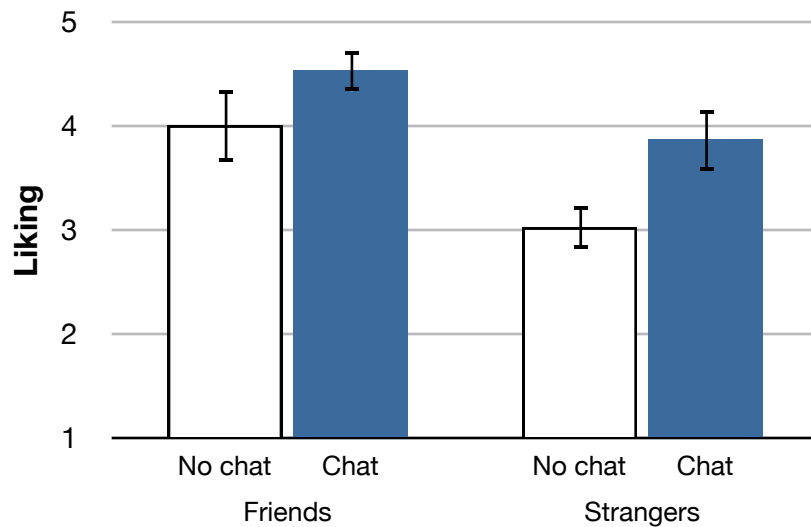


Figure 8-3. Effect of chat on liking of others, between groups of friends and groups of strangers, with and without chat. Error bars represent 95% confidence intervals.

As expected, friends liked each other more than strangers (friends: $M [SD] = 4.4 [.61]$, strangers: $M [SD] = 3.6 [.77]$; $F(1,78) = 32.5, p < .001$). Friends with chat liked each other more than friends without chat (friends/chat: $M [SD] = 4.5 [.54]$, friends/

no chat: $M [SD] = 4.0 [.61]$; $F [1,78] = 6.8, p = .01$). Finally, strangers with chat liked each other more than strangers without chat (strangers/chat: $M [SD] = 3.9 [.77]$, strangers/no chat: $M [SD] = 3.0 [.34]$; $F [1,78] = 15.7, p < .001$).

Figure 8-4 shows the mean values of closeness between groups of friends and groups of strangers, and groups with and without chat. Closeness was measured on a 7-point scale, with 7 being the highest level. Overall, there was a significant main effect of chat on closeness, $F (1,78) = 25.5, p < .001$. Participants with the chat feature felt closer to their other group members more than participants without the chat feature (chat: $M [SD] = 3.6 [1.8]$, no chat: $M [SD] = 2.2 [1.5]$).

As expected, friends felt closer to one another than strangers (friends: $M [SD] = 4.3 [1.7]$, strangers: $M [SD] = 1.9 [.90]$; $F [1,78] = 89.6, p < .001$). Friends with chat felt closer to each other than friends without chat (friends/chat: $M [SD] = 4.9 [1.3]$, friends/no chat: $M [SD] = 3.0 [1.6]$; $F [1,78] = 25.4, p < .001$). Finally, strangers with chat felt closer to each other than strangers without chat (strangers/chat: $M [SD] = 2.1 [.94]$, strangers/no chat: $M [SD] = 1.3 [.44]$; $F [1,78] = 4.68, p = .03$).

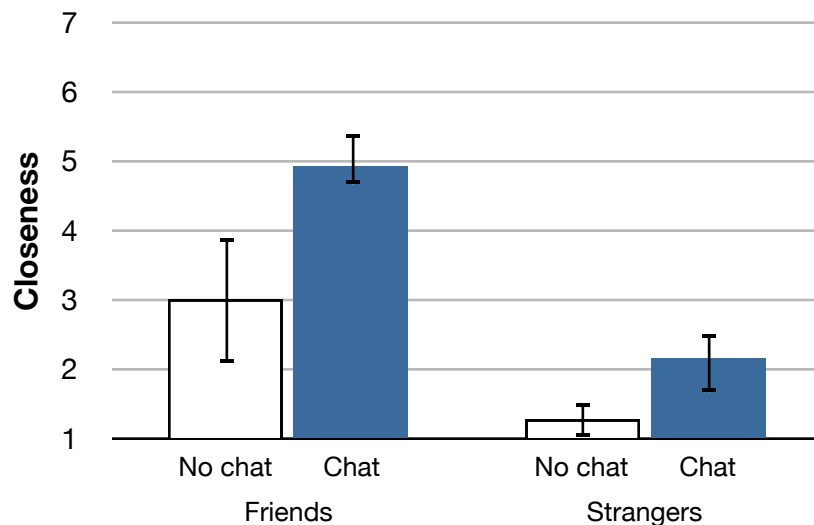


Figure 8-4. Effect of chat on feelings of closeness, between groups of friends and groups of strangers, with and without chat. Error bars represent 95% confidence intervals.

8.7. DISCUSSION

The Cartoon study demonstrates that the tradeoff between chat and distraction may be manageable. Participants had fun in this study, and they enjoyed using the chat feature (RQ 8-1). Indeed, participants who used the chat feature more also enjoyed it more, although as this is a correlation it is unclear if more chat led to more enjoyment of chat, or if more enjoyment of chat led to more chat.

Chat had an interesting effect on ratings of content quality. The ratings of the videos used in this study revealed three underlying groups: videos that were good, videos that were okay, and videos that were poor. For the good and okay videos, participants with and without chat rated them equally. For the poor videos, participants with chat rated them higher than participants without chat. This effect was noticed, qualitatively, in a study of social television by Ducheneaut et al. (2008). In their study, participants mentioned that poor quality movies were a good way to foster social interaction because poor casting, acting, effects, or a lack of important or relevant dialogue provides viewers with opportunities to make comments and jokes. Thus, their qualitative observations along with our quantitative findings strongly support the existence of an “MST3k effect,” whereby making commentary on a bad movie makes the experience more enjoyable¹³.

The second research question asked how distracted participants felt when chatting while watching, whether break periods helped them feel less distracted, and whether they took advantage of the break periods to chat without being distracted by the videos (RQ 8-2). Overall, participants did report feeling distracted by the chat, and there was a significant correlation between the amount a participant chatted and how distracted they felt. Participants with intermissions tended to feel less distracted than participants with the break at the end. Further, participants did use the break periods to chat, although the majority of their chat occurred during the videos (62%). Finally, participants expressed a strong desire for intermission periods between the videos; all participants with intermissions preferred them, and about half of participants with the break at the end would have wanted

¹³ Mystery Science Theater 3000 was a television show produced in the late 1980s and early 1990s. The premise was that a guy, trapped on a spaceship, was forced to watch bad movies by his captors. To maintain his sanity, he built several robot friends that watched the movies with him and made entertaining commentary in the form of riffs, wisecracks, jokes, and skits.

intermissions instead. Thus, we conclude that the intermission periods are a viable strategy for reducing the distraction from chatting while watching videos.

Understanding the topics of chat is the third research question for this study (RQ 8-3). The MovieLens study suggested that viewers could be coaxed to chat about topics other than the movie being watched. The Cartoon study confirms that viewers do go off-topic in their chats, even when those viewers are strangers who do not know each other. This result is promising for collaborative online video communities. It demonstrates that the presence of the videos is not strong enough to discourage off-topic conversation. Off-topic conversation is important for online communities as it is one factor that leads to the formation of bonds between community members (Sassenberg, 2002; Preece & Maloney-Krichmar, 2003). Therefore, this finding supports the sociability argument that relationships can be formed and maintained by watching videos and chatting online.

The final research question for this study asks whether chatting while watching videos has an impact on momentary feelings of liking and closeness to group members (RQ 8-4). We see that it did. For groups of friends – who presumably liked each other before the study – we saw increased levels of liking and closeness when those friends chatted with each other. We saw the same effect for strangers as well. These effects are encouraging because they support the sociability argument: collaborative watching is sociable because people feel the presence of the other viewers when they chat with each other. Although the question of whether collaborative watching can build actual social capital cannot be answered in the context of a one-hour lab study¹⁴, the results from the Cartoon study at least show that the preconditions have been met. In other words, if we did not see an effect of the chat feature on momentary feelings of liking and closeness for friend and stranger groups, then we might conclude that the distraction argument – chatting while watching is too distracting to be enjoyable – is more descriptive of collaborative online video watching.

¹⁴ Strangers did like each other more when they chatted, but that is a far cry from having them become lifelong friends.

8.7.1. LIMITATIONS

This study measured participants' subjective feelings of distraction instead of their objective level of distraction. Thus, it is unclear whether periods of intermission reduced actual distraction as well. Further, the observed reduction in distraction from having intermission periods was marginal. In Chapter 12, I detail a real-world study that shows how online viewers, in control of their own video playback, use periods of intermission to take a break from the videos and chat with each other. Thus, the utility of intermission periods is borne out by these two studies, despite the marginal effect seen in this one.

Another limitation of our measures of distraction is that we are using entertainment videos. Videos of other genres, such as news programs, interviews, or documentaries have more content that viewers need to digest. Thus, chat may be even more distracting for these kind of videos, as there is an overabundance of verbal content. In the next chapter, I discuss a study that compares between videos with verbal content and those with no verbal content. A full treatment of all of the different genres of video is beyond the scope of this dissertation, but is discussed further in Section 17.1.

8.8. SUMMARY AND CONCLUSIONS

- This chapter details a laboratory study in which 15 groups of friends and 15 groups of strangers watched a series of cartoon videos together.
- Participants in 20 groups were able to chat with each other using a text chat feature; participants in 10 groups did not have a chat feature available.
- Participants enjoyed chatting while watching the videos, although they also felt distracted by the chat.
- Intermission periods between videos tended to reduce participants' feelings of distraction from the chat feature. Participants used the intermissions to chat without being distracted by the video, although they did not confine their chat to just the intermission periods.
- Chat had a significant effect on the ratings of poorer videos. Participants with the chat feature rated poorer videos higher than participants without

the chat feature. Thus, we found evidence for an “MST3k effect,” that chatting during poorer content makes that content more enjoyable.

- Participants with the chat feature liked each other more and felt closer to one another than participants without the chat feature. This effect was seen for groups of friends and for groups of strangers.
- The findings in this study provide evidence in support of the sociability argument, that chatting while watching is a sociable experience that can be used to reinforce existing relationships and help create new ones.
- The findings in this study also provide evidence in support of the distraction argument, that people subjectively feel distracted from chatting while watching. However, the extent of this distraction seems minor (in the middle of the 7-point scale), and intermission periods tended to reduce participants’ feelings of distraction.

9.

THE TEXT VS. AUDIO STUDY¹⁵

Thus far in this dissertation, we have considered the combination of online video with a text chat feature. However, there is no technical reason for why the chat feature must be textual. What if remote viewers could converse with one another, using voice chat¹⁶, while watching videos?

There is precedent for using voice chat online. Chapter 2 presented many examples of computer-mediated communication systems that used some form of audio – either through voice chat alone, or combined with video – to enable remote collaborators to communicate with each other. Voice chat may be preferable to text chat because voice is a richer medium – it provides more immediate feedback, it allows for a broader range of cues, and it provides a greater sense of virtual co-presence (Daft & Lengel, 1986; Dennis & Kinney, 1998). When used in the context of online gaming, Williams, Caplan, and Xiong (2007) found that gamers who used voice chat while playing liked and trusted each other more than when they used text chat. Slater, Sadagic, and Schroeder (2000) found that attachment between remote collaborators in a virtual environment increases if members have a sense of virtual co-presence or a subjective feeling of being together. Jensen et al. (2000) found that voice chat led to higher levels of cooperation than text chat in a prisoner’s dilemma task. Another benefit of voice chat is that it makes chatting accessible to people who are uncomfortable with technology or are unable to use computer keyboards. Therefore, using a voice chat feature while watching videos online may further

¹⁵ The results in this chapter appear in part in (Weisz & Kiesler, 2008).

¹⁶ “Voice chat” and “audio chat” are used synonymously.

increase the sociability of the experience, as well as open that experience up to users who might otherwise be excluded.

Voice chat may come with a significant cost in the form of increased distraction while watching a video. The multi-modal model of attention states that two visual (VV) or two auditory (AA) sources of information are highly distracting, but one visual and one auditory (VA) source is not (Wickens & Hollands, 2000). To this point, we have seen that overloading the visual channel by combining video with text chat does result in distraction – both subjective feelings of distraction and objective losses in memory for video content. However, despite this distraction, participants still enjoyed the experience.

With the addition of voice chat, the multi-modal model predicts that viewers will again be distracted; this time, the audio of viewers' conversations will interfere with the audio channel of the video. However, the model is not clear about which situation will result in more distraction: visual overload (VV) or audio overload (AA)? Intuition suggests that audio overload will be more distracting. When someone speaks, his or her auditory speech must be attended to within a few seconds in order to be processed and understood (Wickens & Hollands, 2000). If this speech is garbled or missed, resulting in a broken conversational turn, the only way to recover is by repeating the audio. In textual media, conversational recovery is much easier. The chat history log (shown to not be additionally distracting in Chapter 7) allows viewers to read past messages at their own pace, and to time their reading to the events occurring in the video (e.g., they can read the chat history during downtimes or slow parts). In auditory media, recovery requires playing back the missed audio, either by having the message repeated by one of the other viewers (not necessarily the original speaker), or by having the system cache audio messages and play them back at the viewer's request. Although control over the timing of message playback may help viewers manage their attention, both cases have the same effect: recovering audio means playing it back again over the video content, which may lead to further distraction.

Geerts (2006) examined the use of a voice chat feature among viewers who watched television programs with the AmigoTV system. His participants reported mixed results about the experience. Of seventeen participants, eight reported that it was easy to follow the television program while using the voice chat feature, and five reported that it was difficult. Although some of these difficulties stemmed from the

particular implementation of voice chat (television and voice audio was mixed and presented over the same loudspeaker), the results tend to support using a voice chat feature for collaborative online video.

In the Text vs. Audio study, we perform a quantitative analysis of the distraction present when using a voice chat feature while watching videos online. We also compare voice chat to text chat to determine if auditory overload is significantly worse than visual overload. If voice chat is overly distracting, to the point where it has a significant impact on viewers' enjoyment of the experience, then the multi-modal model is correct in its advice: do not overload the auditory channel. However, if people are able to manage their attention between voice chat and the audio from a video, then voice chat becomes a viable design option for collaborative online video sites that want to provide a more intimate environment for their viewers. For example, the "living room" experience of watching movies on the couch with friends could be replicated online. Indeed, such an application was developed and released by Netflix in 2009, whereby subscribers with an XBox 360 could watch streaming videos from Netflix while chatting with their friends with voice chat. This experience is distinguished by having virtual avatars sit on a virtual couch while watching a movie (Figure 9-1).

We perform one additional comparison in this study, motivated by an observation in the nature of how collaborative online video sites structure the shared watching experience. Some sites, such as UStream.TV, Justin.TV, and the Netflix application above, use streaming video technology that synchronizes viewers with respect to the content they see. In other words, all viewers watch the same video at the same time, modulo network conditions that introduce jitters and delays in video playback. This streaming model of video playback is contrasted with a playlist model, used by sites such as YouTube Streams and Gaia Online. In the playlist model, viewers build their own personal playlist of videos to watch. These videos may be different from the videos being watched by other viewers in the same chat group. The qualitative differences between these two models are discussed further in Section 3.5. In this study, we quantify the effects these models have on the sociability and enjoyment of the collaborative video watching experience.



Figure 9-1. Virtual avatars watch a movie together using the Netflix application on Xbox 360. The real viewers to whom those avatars belong communicate with each other using voice chat.

9.1. RESEARCH QUESTIONS

The research questions for this study are about the different effects of textual and auditory media on distraction and sociability, as well as differences between viewers who watch synchronized content (streaming model) and viewers who watch unsynchronized content (playlist model).

First, we ask whether viewers enjoy using a voice chat feature, whether it helps them feel closer to one another, and what the impact of voice chat is on viewers' distraction.

RQ 9-1: Do viewers enjoy using voice chat? Does it help them feel closer to one another?

RQ 9-2: Is voice chat more distracting to use than text chat?

Next, we ask about differences in the social experience of watching and chatting between viewers who watch the videos at the same time and viewers who watch the videos at different times.

RQ 9-3: How does the social experience differ between viewers watching in a streaming model and viewers watching in a playlist model?

We address these questions in the context of an experimental laboratory study.

9.2. DESIGN

In this study, groups of friends watched videos and chatted with each other in the laboratory. To compare between different chat media, groups with the chat feature were randomly assigned to have either text chat only, voice chat only, or both text and voice chat; this last condition represents a “worst case” situation for distraction, in which both the visual and auditory channels are overloaded. Groups without any chat feature were used as a control. In addition, we compare between groups watching in a streaming model format (same order) and groups watching in a playlist model format (different order). Finally, to compare auditory and visual distraction, two types of videos were shown: videos containing dialog and videos containing no dialogue. Thus, this study uses a 4 x 2 x 2 factorial design with chat media (no chat, text chat, voice chat, or both text and voice) and video order (same order or different order) as between-subjects factors, and video dialogue (dialogue or no dialogue) as a within-subjects factor.

9.3. PARTICIPANTS

Participants were recruited in groups of three friends from the CBDR web site. Forty-eight groups were recruited, for a total of 144 participants. Groups were distributed evenly across the eight combinations of experimental conditions, with six groups per combination of chat media and video order.

The average age of the participants was 23.8 (SD = 7.2) years, and 36% were female. Eighty percent of participants were students (44% graduate, 36% undergraduate). Twenty percent reported other affiliations such as alumni or visiting scholar, or did not list their affiliation. Participants were paid \$15 each for their participation, which took approximately one hour.

9.4. METHOD

Participants in each group watched a series of videos on the computer. Due to the use of voice chat in this study, all participants were seated in separate rooms to more accurately simulate a remote chatting experience. Participants with text chat could type messages to one another using web-based chat software. Participants with voice chat could speak to one another using headsets we provided. Participants were given time at the beginning of the study to test the voice chat feature, adjust their audio to comfortable levels, and ensure that they could hear each other properly. In each chat condition, participants were told that they could chat with other participants at any time during the study, about any topic. Participants were only told of the availability of the chat feature, not that it was a mandatory requirement of their participation.

In this study, we used a slightly different software configuration from that used in the Cartoon study. Here, the video and text chat were placed on a web page in a fixed position, preventing participants from rearranging the positions of chat and video windows on screen. This fixed placement was done to minimize the ability of participants to distract themselves by rearranging windows, as well as to eliminate a potentially confounding independent variable of window position and size. A picture of the software setup is shown in Figure 9-2. For groups not assigned to have any chat feature, and for groups assigned to have only the voice chat feature, the text chat box was not displayed in the browser window. For groups with the voice chat feature, the TeamSpeak software was run in the background. It was configured to use the voice-on activation feature so participants could speak to each other without having to push a separate button.

Participants watched the eight Text vs. Audio study videos discussed in Section 5.6. Twenty-seven participants (18.7%) reported previously seeing at least one of the videos. Each video lasted between three and a half and six minutes.

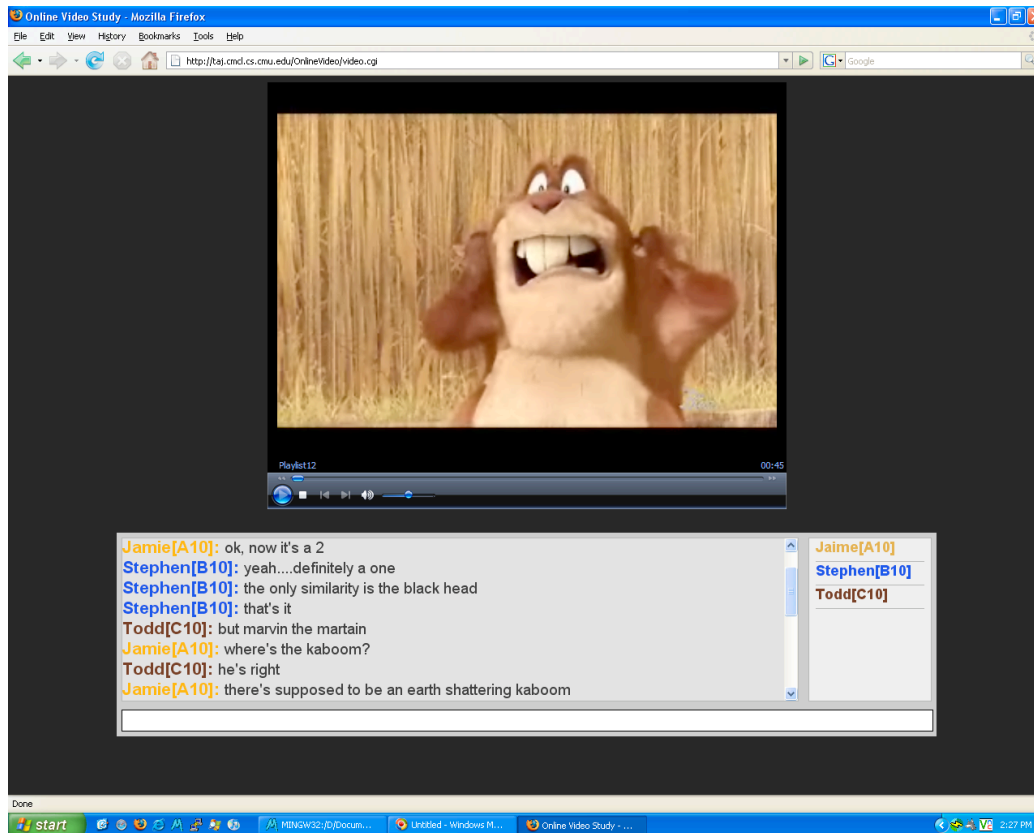


Figure 9-2. Screenshot from the Text vs. Audio study. The text chat feature was not displayed for participants in the no-chat and voice-chat conditions. Participants with voice chat or both text and voice chat could speak to each other over their headsets.

Participants in the same-order condition watched the videos in the same order; their video playback was synchronized such that they all saw the same video at the same time. Participants in the different-order condition watched the videos in a different, randomized order from each other. Since it was possible that purely randomized orderings would result in participants watching the same first video, and thus begin with a synchronized experience, we ensured that the first video seen by each participant was different from each other. This randomization kept the different-order situation realistic to playlist model sites, as newcomers are likely to start off watching a video different than that of other viewers. Figure 9-3 depicts the difference in video playback between same-order and different-order groups. Note that as the length of each video differed, transitions between each successive video were not temporally synchronized for each participant. Again, this feature is reflective of the playlist model.

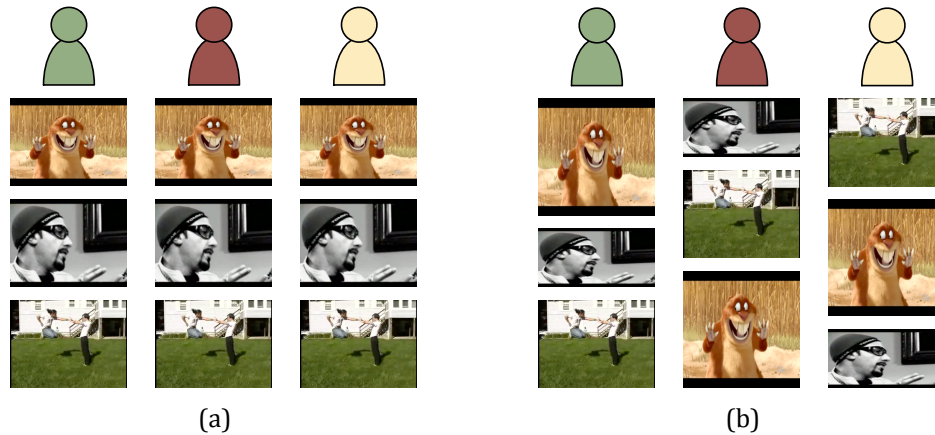


Figure 9-3. (a) Participants watch the same videos at the same time. Each participant’s video is synchronized with the other members of their group. (b) Participants watch the videos in a different, random order from each other. The initial video for each participant is always different. Transitions between successive videos do not occur at the same points in time as the videos are of different lengths.

9.5. MEASURES

All participants rated each video immediately after it had finished, to avoid difficulties in recall. These ratings were made on the same 5-point scale used in the Cartoon study. All participants also completed a final survey at asked about their experience and the chat feature(s) they had (where appropriate).

Enjoyment of the study was measured by asking participants “How would you rate the experience of participating in this study” on an open-ended 7-point scale, anchored by “Very boring” (1) and “Very fun” (7). Enjoyment of the chat was measured using the same chat enjoyment scale used in the Cartoon study. This scale remained reliable ($\alpha = .78$).

As a manipulation check, participants were asked which chat features they had available to them, as well as whether they watched the videos in the same order as their friends or in a different order. All participants correctly reported the chat features available to them. Eighteen participants (12.5%) were unsure of their video order condition; 10 were in the different order condition and 8 were in the same order condition. No participants thought they were in the opposite video order condition. Thus, we are confident that our manipulations took effect, although we

note that in a real-world situation, an explicit signal may be required to inform viewers when they are or are not synchronized.

Participants were also asked how comfortable they were using the chat features provided to them. Two questions were used to assess media comfort, listed in Appendix B ($\alpha = .71$). Participants with both text and voice chat were asked the media comfort questions twice; once for each medium.

Participants with any chat feature were asked how distracted from chat they felt (listed in Appendix B). To objectively determine whether voice chat was more distracting than text chat, we also used video memory measures similar to those used in the Chat Distraction study (Chapter 7). In order to not disturb participants' social experience, the memory questions about the videos were all asked as part of the final survey. To prevent an overabundance of questions on the final survey, only one question was asked about each of the eight videos. These questions asked about things participants heard and things participants saw in the videos. A few sample questions asked about these videos are given in Appendix B.

To measure sociability, we used the same liking and closeness scales as in the Cartoon study. The liking scale was administered as part of the final survey, and remained reliable ($\alpha = .85$). For the closeness scale, we attempted to improve on our measure of closeness by taking measurements at two times: first, before the study began (denoted C_1), and second, after all of the videos finished playing (denoted C_2). Since each group consisted of three people, each participants' ratings of closeness at these times represents their average rating for their other two friends.

Finally, we added measure of engagement to this study. Engagement is a measure of how much attention participants paid to watching the video and chatting, as opposed to other activities such as looking around the room or falling asleep. We operationalized engagement by giving participants pretzels to eat while watching (a potentially distractive activity), and measuring how much they ate. We hypothesized that participants who were highly engaged in the experience of watching and chatting would ignore and/or forget about the pretzels. Thus, we defined engagement as the percentage of pretzels a participant did not eat, as fully engaged participants should not eat any pretzels, resulting in an engagement score of 1.0. We note that our operationalization of engagement is, of course, confounded with the uncontrolled variables of hunger and dieting.

9.6. RESULTS

Unless otherwise specified, the primary method of analysis for the outcome variables in this study is an analysis of variance (ANOVA). In the ANOVA models, the explanatory variables are chat type (no chat, text chat, voice chat, and both text and voice chat) and video order (same order or different order). Group ID is also included in the model as a random effect to control for within-group variance. Contrast testing is used to compare between specific conditions (e.g., text chat vs. voice chat), as well as between groups that had the chat feature versus groups that did not.

To compare the content of chat between the different conditions, we logged the text chats and we recorded and transcribed the voice chats. For the groups with voice chat, we transcribed their speech such that one thought or phrase corresponded to one line in the transcript. For example, when two speakers alternated in speaking, each alternating turn was a separate line in the transcription. When one person spoke, paused for a moment, and then spoke again, the pause was considered to be the beginning of a new conversational turn, and was placed on a separate line. We used two seconds as a rough guideline for the length of these pauses, but also considered whether the content after the pause was related to what was previously said. For example, “Are they fishing for something? (*pause*) It looks like they have nets” was kept together, because the statement gives the reason for the question.

One final note about transcribing the audio logs is that many participants laughed out loud in the study. This out-loud laughter was coded as “[laughter]” in the text transcriptions, and is counted as a single word when analyzing word counts.

9.6.1. ENJOYMENT, CLOSENESS, AND CHAT MEDIA

Study enjoyment. Overall, participants had fun participating in this study ($M [SD] = 5.0 [1.3]$ of 7). Study enjoyment was not significantly different between the different chat groups ($F [3,40] = .65, p = n.s.$). Participants watching the videos in a different order tended to enjoy the study more than participants watching in the same order (different order: $M [SD] = 5.2 [1.2]$, same order: $M [SD] = 4.7 [1.4]$; $F [1,40] = 3.3, p = .07$). The interaction between video order and chat type on enjoyment was not significant, $F (3,40) = .43, p = n.s.$

Chat enjoyment. The first research question is whether viewers enjoy using voice chat while watching the videos (RQ 9-1). Overall, participants with the chat features enjoyed using them ($M [SD] = 3.8 [.86]$ of 5). There was no difference in chat enjoyment between groups with only text chat and only voice chat (text chat: $M [SD] = 3.9 [.87]$, voice chat: $M [SD] = 3.7 [.88]$; $F [1,30] = .38$, $p = n.s.$).

Participants watching the videos in the same order enjoyed using the chat feature more than participants watching the videos in a different order (same order: $M [SD] = 4.0 [.79]$, different order: $M [SD] = 3.6 [.88]$; $F [1,30] = 4.1$, $p = .05$).

Liking and closeness. Participants with voice chat reported a mean liking score of 4.0 ($SD = .98$) of 5. Participants with text chat reported a mean liking score of 4.3 ($SD = .65$). This difference was not significant, $F (1,40) = 1.96$, $p = .17$. Therefore, participants with voice chat did not like each other any more or less than participants with text chat.

Participants watching the videos in the same order reported a mean liking score of 4.1 ($SD = .70$). Participants watching the videos in a different order from each other reported a mean liking score of 4.2 ($SD = .74$). This difference was not significant, $F (1,40) = .002$, $p = n.s.$ Therefore, participants watching in the same order did not like each other any more or less than participants watching in a different order.

In comparing the closeness measures between time 1 and time 2, we found that they were significantly correlated ($r = .91$, $p < .001$). In addition, the mean difference between the two closeness measures was .05 ($SD = .59$). Thus, there may have been a memory bias in this measure whereby participants simply remembered their closeness ratings from time 1 and repeated them for time 2. Therefore, to compare closeness between chat groups, we added C_1 to the model as a covariate to control for the initial closeness ratings. Since C_1 is a nominal measure, this analysis is an analysis of covariance (ANCOVA). With this model, we found that participants using only text chat did not differ from participants using only voice chat in their feelings of closeness to each other, $F (1,39.0) = 1.6$, $p = n.s.$ Therefore, although voice chat is a more intimate medium than text, we do not find evidence that it increased participants' momentary feelings of closeness to one another.

We also do not see a difference in closeness between participants watching the videos in the same order and participants watching the videos in a different order, $F (1,39.0) = 1.33$, $p = n.s.$

Media comfort. We asked participants two questions about their comfort using the different chat media on a 5-point Likert scale. Participants with only text chat had a mean media comfort score of 3.7 (SD = .81), and participants with only voice chat had a mean media comfort score of 3.6 (SD = .93). This difference was not significant, $F(1,20) = .05$, $p = n.s.$

Participants with both text and voice chat were asked the questions on the media comfort scale twice: once for text and once for voice. They reported a mean text comfort of 3.7 (SD = .8) and a mean audio comfort of 3.6 (SD = .9). Thus, we conclude that participants generally felt comfortable with the chat feature(s) they had.

Media preferences. Although we did not see a significant effect of voice chat on enjoyment of chat or feelings of closeness, these results do not speak to participants' preferences for each type of media. In the final survey, we asked participants which chat medium they would have preferred to use in the study: no chat, text chat only, voice chat only, or both text and voice chat. The responses across the different chat media conditions are shown in Figure 9-4.

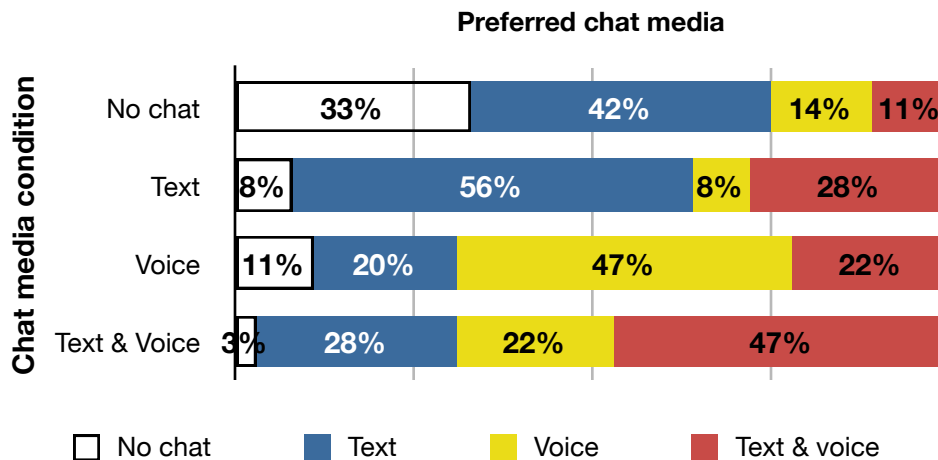


Figure 9-4. Preferences for different chat media by chat media condition. Each horizontal strip shows the distribution of media preferences for the participants in the specified condition.

Two-thirds of participants without chat expressed a desire to have some chat feature while watching the videos; 42% expressed an interest in text chat, and 25% expressed an interest in voice chat, either alone or in conjunction with text chat. Thus, many participants without chat felt they would have wanted to have a text chat feature while watching the videos.

For the participants who had the text chat feature, the majority expressed satisfaction with their medium: 56% expressed that they preferred the text chat. Thirty-six percent of participants with text chat felt that they would have wanted voice chat, either alone or in conjunction with text chat. Thus, we see that the majority of participants with text chat preferred it, although there were some participants interested in voice chat.

An interesting effect is seen when considering participants who were exposed to voice chat, either alone or in conjunction with text chat. For participants in each of these conditions, a majority (69% in both cases) expressed a preference for voice chat, either alone or in conjunction with text chat. Further, for the participants that had voice chat, 47% preferred keeping just the voice portion of the chat.

The media preferences for participants with a chat feature a bias toward the chat media they experienced in the study. Participants with text chat generally preferred having text chat, and participants with voice chat generally preferred having voice chat. This finding is telling for several reasons. First, it shows that people's intuition is that text chat would be more enjoyable to use while watching videos (the participants without chat also expressed this opinion). However, once participants were exposed to the voice chat feature while watching videos, their preferences shifted toward options that included voice chat.

In each chat condition there was a small group of participants that expressed a preference to not chat (3-11%). This finding mirrors the comments made by participants in the MovieLens study, that they are just not interested in chatting while watching a video.

Interestingly, participants' preferences changed when they were asked to speculate on chatting while watching videos with strangers. In this case, participants expressed a strong overall desire to use text chat (62%) over voice chat (22%) when watching with strangers (the last 16% reported not wanting chat). Even among those who had voice chat in the study, 53% reported wanting to use text chat with strangers. Several reasons were given for why participants preferred text chat with strangers (all quotes are reproduced in their original form).

"I'm shy around strangers" (TA123)

"[It's] less intimidating" (TA7)

“texting with people i don't know is more comfortable” (TA30)

“Audio chat is more suited with friends. With strangers there can be some awkwardness initially” (TA141)

Media and content. To further gauge participants' preferences for different chat media, and their feelings of whether text or voice chat are better suited for video content having differing amounts of dialogue, participants were asked if they had a preference for text or voice chat with these different types of videos. Overall, about half of all participants, and 75% of participants with both text and voice chat, felt that different chat media were suited for videos with differing amounts of dialogue. Some participants were sensitive to the auditory distraction.

“If there is no talking in the video, chatting is ok, but when there was talking, the texting was more appropriate” (TA3, both text and voice)

“It's better to text if the movie relies heavily on dialog and is interesting” (TA48, both text and voice)

“I think using an audio chat would totally disrupt the attention given to watching a video” (TA85, text chat)

Other participants felt the asynchrony of the text chat was beneficial, and that the text chat was less distracting.

“Text chat is always less distractive since I can answer it anytime I would like to, audio chat requires me to answer right away” (TA18, both text and voice)

“Sometimes I want to focus on the video without being distracted, and then I choose text chat. Otherwise, I choose audio.” (TA24, both text and voice)

Some participants felt that audio was advantageous because of its immediacy and ease of use.

“Audio chat allows you to voice your immediate reaction” (TA95, voice chat)

“Audio chat helps to get the message quick and fast” (TA57, both text and voice)

“Well I don't think chatting using text will be possible because I will not be able to concentrate on the video. And typing will naturally be slower than speaking!” (TA80, voice chat)

Finally, for some participants, chat media didn't matter.

“To me it doesn't matter as long as you can talk to someone” (TA21, text chat)

“text vs audio is matter of personal preference and doesn't depend on content seen” (TA134, voice chat)

9.6.2. DISTRACTION

Feelings of distraction. The second research question asks whether distraction is greater when overloading the auditory channel (with voice chat) or the visual channel (with text chat). Participants with only voice chat reported a mean distraction of 4.2 (SD = 1.9) on the 7-point self-reported distraction scale. Participants with only text chat reported a mean distraction of 3.6 (SD = 1.6). The contrast between these two groups in the ANOVA model showed no significant difference, $F [1,30] = 2.0, p = .16$. Therefore, participants with only voice chat did not feel any more or less distracted than participants with only text chat. In addition, participants with both text and voice chat reported a mean distraction of 3.7 (SD = 1.6). The ANOVA model showed no main effect of chat media on distraction, $F (2,30) = 1.2, p = \text{n.s.}$ Thus, participants with both text and voice chat were not significantly more distracted than groups with only one type of chat.

Memory. To quantify the degree to which participants were distracted, we asked them questions about the videos they watched. Of the eight memory questions, participants with voice chat answered an average of 5.8 (SD = 1.3) questions correctly. Participants with text chat answered an average of 5.4 (SD = 1.6) questions correctly. The contrast between these two groups shows that this was not a significant difference, $F (1,40) = 1.7, p = .20$. Further, participants with both text and voice chat answered an average of 5.4 (SD = 1.3) questions correct; a contrast between these participants and participants with either text or voice chat showed

no significant overload from simultaneously having both chat media, $F(1,40) = .57$, $p = n.s.$

Participants without chat answered an average of 6.7 ($SD = 1.2$) questions correct. The contrast between participants with any form of chat and participants without chat showed a significant difference, $F(1,40) = 16.8$, $p < .001$. Thus, the main distractor while watching videos is the presence of the chat feature itself, rather than the specific medium embodied by the chat.

To determine which type of overload is more distracting – visual (VV) or auditory (AA) – participants watched two types of videos: those with significant verbal content, and those with no verbal content. If overloading the auditory channel is more distracting than overloading the visual channel, we would expect participants with voice chat to remember more about videos without dialogue than participants with text chat, and participants with text chat to remember more about videos with dialogue than participants with voice chat. Neither of these effects are seen in the data. For the videos with dialogue, participants with text chat remembered as much as participants with voice chat, $F(1,40) = .61$, $p = n.s.$ For videos without dialogue, participants with text chat remembered as much as participants with voice chat, $F(1,40) = 1.3$, $p = n.s.$

Engagement. Our measure of engagement was designed to measure the degree to which participants “lost themselves” in the chatting and watching experience. It was operationalized by providing a bowl of pretzels to participants, and measuring how many pretzels were consumed. Engagement ranged from 0.0 to 1.0. Higher levels of engagement represents fewer pretzels consumed. Participants with voice chat had a mean engagement score of .59 ($SD = .37$), and participants with text chat had a mean engagement score of .28 ($SD = .36$). This difference was significant, $F(1,40) = 6.4$, $p = .01$. Therefore, participants with voice chat showed higher levels of engagement by eating fewer pretzels than participants with text chat. One explanation for this result is simply that participants with voice chat were being polite to their friends by not eating crunchy, noisy pretzels.

9.6.3. CHAT USAGE AND CONTENT

Usage of the chat feature. One explanation for why participants with voice chat did not feel more distracted than participants with text chat is that they simply did not

use the chat feature, or did not use it as much as participants with text chat. Thus, we compare the amount of chat between groups with text chat and groups with voice chat. In making this comparison, we use word counts rather than line counts because words are a more accurate measure of how much participants in each group chatted. Word counts are also not affected by our transcription method.

Groups with only text chat typed an average of 1,020 (SD = 379) words in their chats. Groups with only voice chat spoke an average of 2,202 (SD = 1,379) words in their chats. This difference was significant, $F(1,30) = 8.15, p < .01$. Thus, groups with only voice chat spoke twice as many words as groups with only text chat typed. In addition, groups with both text and voice chat mainly used voice chat; they typed an average of 401 (SD = 370) words and spoke an average of 1,405 (SD = 729) words. This difference was also significant, $F(1,20) = 15.5, p < .001$. Therefore, we see that groups with a voice chat feature not only used it, they used it quite a lot.

As for the effect of video order on usage of the chat feature, same-order groups uttered about 1747.0 (SD = 813.8) words while watching, and different-order groups uttered about 1560.5 (SD = 1303.6) words. This difference was not significant, $F(1,30) = .3, p = \text{n.s.}$ Therefore, participants spoke just as much when they were watching the videos in the same order as when they watched them in a different order.

Chat content. We coded the chat logs in this study to gain insight into how the topics of conversation differed between groups using different chat media, and between groups watching the videos in the same order or in a different order. We based our coding scheme on the one used in the Cartoon study, with two additions. As some of the audio in the voice chats was unintelligible, we added a category for “Unintelligible” chat. A few foreign language statements were present in our corpus as well, and these were also classified as unintelligible. We also added a code for “Coordination,” as there were a significant number of comments and questions about which videos participants’ were watching. The distribution of the amount of chat in each coding category is shown in Table 9-1.

The entire chat corpus contained 10,813 lines of chat. A subset of this (869 lines, 8%) was used for a reliability analysis. Two independent coders coded this subset of data and achieved a Cohen’s κ of .71. This is an adequate level of reliability for analysis.

Table 9-1. Distribution of chat content across coding categories. The overall distribution is given, as well as distributions for groups based on video order and chat media.

Category	All groups Total	Video order		Chat media	
		Same order	Different order	Text only	Voice only
Videos	36.1	45.3	25.0	43.0	37.1
Study	17.8	15.6	20.4	17.3	20.2
Laughter	16.9	17.5	16.2	7.3	12.0
Evaluations	12.0	10.5	13.8	13.6	11.9
Personal	9.9	7.5	12.8	12.8	10.6
Coordination	5.4	1.4	10.3	4.8	5.5
Greetings & partings	1.4	1.3	1.4	1.1	1.5
Unintelligible	0.5	0.9	0.1	0.3	1.1

Overall, the videos were the most popular topic of conversation (36%). This finding is consistent with the amount of cartoon-related chat in the Cartoon study (~42%). Participants also spoke a bit more about this study (~18%) compared to participants in the Cartoon study (~13%). Participants in this study laughed more than participants in the Cartoon study (~17% vs. ~7%), although this difference is a result of having voice chat; participants in this study with only with text chat laughed a comparable amount to participants in the Cartoon study (~7%). Personal topics were less popular in this study than the Cartoon study (~10% vs. ~23%).

Video order had an effect on topics of conversation. Participants watching the same videos at the same time focused more of their conversation on those videos (45.3%) than participants watching the videos in a different order (25.0%). Since there was no significant difference in the amount that participants in these two groups chatted, the decrease in video-related chat for different order groups resulted in increases in chat for several other categories. These increases included more study chat (~16% vs. ~20%), more evaluative chat (~10% vs. ~14%), more chat about personal topics (~7% vs. ~13%), and more coordination chat (~1% vs. ~10%).

Chat media also had an effect on topics of conversation. Participants with voice chat spent less of their chat on the videos (~37% vs. 43%) and more of their chat on laughter than participants with text chat (12% vs ~7%).

9.7. DISCUSSION

The Text vs. Audio study makes a strong case for the inclusion of voice chat in collaborative online video experiences. Participants with voice chat used it, enjoyed using it, and expressed a strong preference for it, either alone or in conjunction with text chat (RQ 9-1). Even the participants who had both text and voice chat expressed a strong preference for voice chat, again either alone or in conjunction with text chat. This preference for voice chat is contrary to the preference for text chat found by Scholl, McCarthy, and Harr (2006). In their study, they found that 60% of their participants preferred using a text chat feature while collaborating on a course project, whereas only 40% preferred using the voice chat feature. One explanation for this discrepancy is that the participants in the Scholl et al. study chatted in a work situation, whereas the participants in the Text vs. Audio study chatted in an entertainment situation. Our findings do mirror those of Geerts (2006) and Harboe et al. (2008a). Geerts (2006) found that groups of friends and family liked using voice chat while watching television, even though they felt it was distracting. Harboe et al. (2008a) found that groups of friends and family enjoyed using voice chat with each other while watching television together in different remote locations.

Interestingly, participants only expressed a strong preference for voice chat after they had been exposed to it. Participants with text chat expressed a strong preference for text chat, and participants without chat expressed roughly similar preferences for either text chat or no chat. However, these preferences were reflective only of chatting with one's friends; when participants were asked which media they would prefer when chatting with strangers, the majority indicated that text chat would be more appropriate because it would make them feel more comfortable. This sentiment is understandable from Walther's hyperpersonal model of computer-mediated communication: text chat is less intimate and less immediate, and enables a higher degree of selective self-presentation because chatters have the opportunity to compose and edit their messages before broadcasting them (Walther, 1996; Walther, 2007). Voice chat does not afford the same degree of editing, as once something is said it cannot be taken back.

These findings indicate that there may be overall resistance to the adoption of a voice chat feature in collaborative online video sites. Friends may not adopt a voice chat feature because of their feelings that it would be too distracting to use. Strangers may not adopt a voice chat feature because of the awkwardness of using a

more intimate medium. In both cases, community leaders and early adopters may be helpful in introducing community members to a voice feature. In addition, the voice feature can be scaffolded through the use of a text chat feature. This way, viewers can “graduate” to voice chat when they feel comfortable and are ready. Providing both features also allows viewers to choose the less-distractive medium depending on the content they are watching.

Another argument that favors the combination of textual and auditory media is that text chats can support more users than voice chats. Löber, Schwabe, and Grimm (2007) found that smaller groups of four people were more satisfied with voice chat than larger groups of seven people. This finding, combined with the preference for text chat with strangers, suggests a design for collaborative online video watching that provides audio for small groups of friends and text for larger groups of strangers.

Because audio is a richer medium than text, we expected participants using voice chat would have increased feelings of closeness to one another (RQ 9-1). Although this effect was not seen in this study, participants’ preferences for voice chat speak more to its value. In addition, the increased amount of laughter in the voice chat groups suggests that participants using voice chat did enjoy chatting out loud with their friends while watching the videos.

The multi-modal model of human attention explains that simultaneously processing two visual (VV) or auditory (AA) sources is extraordinarily difficult for people, and much information is lost when attending to both sources. In this study, we questioned whether an overloaded auditory channel was more distracting for viewers than an overloaded visual channel (RQ 9-2). Intuition suggests that it should be; text chat does not demand one’s attention the same way that voice chat does because text can be ignored or deferred at one’s leisure. Voices must be attended to immediately, or the contents of the utterance will be lost after a few seconds (Moray, 1969). However, intuition was not supported by the results of this study. Participants with only voice chat did not report feeling more distracted than participants with only text chat. They also chatted twice as much as participants with text chat. Although it is easier to speak than type, we would not have seen this outcome if the participants with voice chat found it too distracting to use. Using memory of the videos as a proxy for the degree to which participants were distracted, we did not see a difference in recall among the different chat conditions.

The only difference in recall was between participants who had a chat feature and those who did not; thus, the main distractor is the chat feature itself, and not the specific medium through which chat is conducted.

In addition, we expected participants with both voice and text chat to gravitate toward the less-distracting medium since they had control over which medium they used. If voice chat was more distracting than text chat, these participants should have used the text chat feature more. Thus, the usage of each chat medium is a behavioral measure for chat media preference, and participants with both text and voice chat again expressed a strong preference for voice chat. These participants used the voice feature more than thrice as much as the text feature ($M = 401$ words typed vs. $M = 1,405$ words spoken). We must therefore conclude that voice chat ought to be seriously considered as a feature for collaborative online video, as participants enjoyed using it, and it did not significantly distract them more than a text chat feature.

In the comparison between the two models of video playback – streaming and playlist – we did not see any differences in the social experience (RQ 9-3). Participants felt just as close when watching different videos and they chatted with each other just as much. There were a few key differences in the topics of their chat. Participants watching the same videos at the same time were more on-topic than participants watching videos in a different order. Participants watching in a different order focused more on other sources of common ground, such as the study in which they were participating and personal topics. This finding is interesting, as it suggests that collaborative online video sites may be able to shape the social dynamics of the community by influencing the content of viewers' conversations. Community members can experience two forms of attachment to the community: identity-based attachment, for example by being among other Steelers fans, and bond-based attachment, by having friends in the community (Prentice, Miller, & Lightdale, 1994; Sassenberg, 2002; Ren, Kraut, & Kiesler, 2007). A community that wishes to promote common identity may have viewers watch videos in a streaming model. A community that wishes to promote bonding among members may allow viewers to view different videos from each other (in a playlist model), which reduces the amount of explicit, video-based common ground viewers have with each other. The results from this study suggest that this change will result in an increase in personal topics, and therefore, bonding. Confederates or moderators can also be employed to achieve the same effect, by steering conversations toward personal topics.

Having viewers watch different videos from each other does have other effects on their topics of conversation. About 10% of the chat in different-order groups was spent on coordination – asking each other what they were watching. Coordination chat was almost non-existent in same-order groups. Whether this chat wasted opportunities for chat on other, more relevant topics is unclear. Although it may be the case that time spent coordinating could have been better spent on chatting about other things, coordination may be a valuable way for viewers to segue into talking about their videos, and for recommending videos to watch. In systems with much larger playlists of videos, it is not guaranteed that all viewers will watch all of the same videos in the same session. Therefore, querying other viewers about the videos they are watching may be useful in receiving recommendations for what to watch (or what to avoid).

9.7.1. LIMITATIONS

We may have inadvertently introduced a memory bias in the closeness data by measuring closeness twice on a self-report scale. Because there was so little variation between the two measures – the mean difference between the two measures was only .05 – participants may have simply remembered their answer from time 1 and reported that value at time 2 without re-evaluating their feelings. To improve on this limitation, future studies should consider other measures of closeness beyond the Inclusion of Other in Self scale (Aron et al., 1991).

Our measure of engagement showed that participants with voice chat were more engaged than participants with text chat because they ate fewer pretzels. However, concluding that participants with voice chat were actually more engaged than participants with text chat depends on whether (not) eating pretzels was truly a measure of how engaged one was in watching. Further, even if we believe that it was, one explanation for why participants with text chat ate more pretzels than participants with voice chat is simply that eating pretzels makes a lot of noise, and this noise may have been transmitted over the audio channel. Indeed, some participants made comments to each other to this effect.

“mila i can hear you chewing” (TA46)

“I can hear you chew on your pretzel honey” (TA53)

“Can you hear me crunching?” (TA69)

Participants with voice chat may have just been polite, instead of being engaged, by not eating their pretzels. Therefore, we cannot conclude that participants with voice chat were more engaged. For future studies, we recommend developing other creative, behavioral measures of engagement that do not interact with social norms.

9.8. SUMMARY AND CONCLUSIONS

- This chapter details a laboratory study in which 48 groups of friends watched a series of videos while chatting with each other. This study compared two chat options: a text chat feature and a voice chat feature. It also compared two models of video playback: the streaming model (synchronized content) and the playlist model (unsynchronized content).
- Text chat was preferred to voice chat by participants who had only text chat, and voice chat was preferred to text chat by participants who had only voice chat. The bias against voice chat by participants who did not experience it suggests that collaborative online video sites may face adoption difficulties for a voice chat feature. Providing opportunities for viewers to use a voice chat feature in a sandboxed setting (i.e., during a video for which the viewer would not mind being distracted) may help viewers overcome this bias.
- Behaviorally, participants with both text and voice chat spoke more than three times the number of words than they typed. Participants with only voice chat spoke twice as many words as participants with text chat.
- Voice chat was no more distracting than text chat. The presence of any chat feature was the primary distractor, and not the specific medium used for the chat.
- Chat media did not greatly shift the topics of conversation, though participants with only voice chat laughed more than participants with only text chat. Applications that utilize the occurrences of laughter during a video, such as those discussed in Chapter 16, may wish to provide a voice chat feature to viewers in order to elicit more laughter.
- Watching the videos in a different order (playlist model) did not adversely affect the amount participants chatted with each other. It did alter the

distribution of topics of conversation: participants chatted less about the content of the videos, and they chatted more about figuring out which video their friends were watching.

- Participants watching the videos in a different order chatted more about personal topics, suggesting that collaborative online video sites that wish to promote bonding amongst their members have viewers watch different content from each other.

10.

OVERALL DISCUSSION OF THE EMPIRICAL STUDIES

The studies discussed in Part II make a case for the inclusion of a real-time chat feature for online video. These studies found evidence in support of the sociability argument: collaborative watching is able to provide a fun and sociable experience, and participants who chatted with each other felt each others' presence. These studies also provided evidence in support of the distraction argument: viewers who multitasked between watching a video and chatting with others felt distracted from doing so, and remembered less about the videos they were watching. Thus, both arguments are descriptive of collaborative viewing.

This chapter summarizes the main findings from the studies in Part II and discusses the broader implications of collaborative watching in the context of online communities. It concludes with a recommendation that online video sites provide social interaction features for their viewers and discusses how these features can be scaffolded to help viewers manage their attention during the course of watching a video. This recommendation serves as a segue in to Part III of this dissertation, which presents concrete designs and design guidelines for the creation of large-scale collaborative online video experiences.

10.1. SOCIABILITY

The sociability argument states that chatting while watching is beneficial for viewers because it can create and reinforce social relationships. In the studies, participants used and generally enjoyed using the chat feature(s) provided to them. In the

Cartoon study, use of the text chat feature was associated with increases in participants' momentary feelings of liking of each other and feelings of closeness to one another. This effect held both for groups of friends, who presumably liked each other and felt close to one another before the study, as well as groups of strangers who had not met each other prior to the study.

The increases in liking and closeness for groups of friends is important because it demonstrates that collaborative watching can be used to help friends maintain their social networks across great distances. For example, students making the transition between high school and college face the possibility of losing friendships because of moving away from each other. Prior research has shown that communicating online will slow the decline in feelings of psychological closeness for these students (Cummings, Lee, & Kraut, 2004). Chatting while watching movies online is one activity that can be used as an excuse for communication, and thus, could help them maintain their relationships when they lose the ability to make physical contact.

Chatting while watching videos can also be used to create new friendships. The increases in liking and closeness for groups of strangers reinforce prior research results that people bond in online communities through frequent communication and interaction (Ren, Kraut, & Kiesler, 2007; McKenna, Green, & Gleason, 2002). Although the presence of the videos was likely not a factor in the increased feelings of liking and closeness, they do serve as a "social glue" that brings people together and gives them an excuse to have a conversation.

Newcomers are the lifeblood of any online community, and attracting newcomers and engaging them in community activities is of paramount importance for increasing the value and lifespan of that community (Butler, 1999; Ren, Kraut, & Kiesler, 2007). For online video communities, fun activities like watching videos generally attract newcomers; in fact, over half of adult Internet users in the US watch videos online (Madden, 2007), representing a very large potential viewership. Combining a chat feature with the video player helps bootstrap interactions between newcomers and existing members because the videos give them something to talk about. Involving newcomers in conversation with other community members has been shown to increase their commitment and involvement with the community (Arguello et al., 2006; Lampe & Johnston, 2005).

One concern about using videos to bootstrap conversation is that the content of the videos may be overly suggestive of conversational topics as to discourage other topics. In an online community that wishes to promote bonding amongst its members, discouraging personal topics is counter-productive (Ren, Kraut, & Kiesler, 2007). In the empirical studies, we found that although much of the chat was focused on the videos themselves, personal topics did emerge naturally as part of the conversation between friends (Cartoon & Text vs. Audio studies) and strangers (Cartoon study). In addition, personal topics could be broached with a little encouragement from a moderator or confederate (MovieLens study). Off-topic conversation is key for creating interpersonal relationships online (Sassenberg, 2002; Ren, Kraut, & Kiesler, 2007). Because personal chat did occur while watching the videos, both among friends and among strangers, and in cases in which participants watched videos outside of the laboratory, we found more support for the argument that collaborative online video watching can help viewers build social capital with each other.

Chat media had no quantifiable effect on feelings of sociability. Participants using voice chat in the Text vs. Audio study reported equivalent levels of feelings of liking and closeness to their friends as participants with text chat. This finding seems to contradict the belief that audio – by virtue of being a richer medium of communication – provides a more intimate setting and greater feelings of awareness and virtual co-presence (Daft & Lengel, 1986; Slater, Sadagic, & Schroeder, 2000). However, another explanation is that our measures of sociability may not have been sensitive enough, as they relied on self-report. Other researchers have found increases in liking and trust (Williams, Caplan, & Xiong, 2007) and increases in cooperation and trust (Jensen et al., 2000) among people who used a voice chat feature. These results, coupled with the strong preferences for voice chat from the participants who used it, support its use for collaborative online video.

10.2. DISTRACTION

Congruent with the predictions made by the multi-modal model of attention (Wickens & Hollands, 2000), we did find that chatting while watching is a distracting activity. Watching a video requires one's visual and auditory attention, and adding a chat feature results in an overload of one (or both) of these information-processing centers. In the studies, participants reported feeling the effects of this overload, by

complaining of distraction in the MovieLens study, and by reporting their distraction on the self-report distraction scale in the Cartoon and Text vs. Audio studies.

We investigated distraction further by seeing if it had an impact on participants' ability to follow the events in each video and recall them later. Both the Chat Distraction and the Text vs. Audio studies showed that when participants had any chat feature – either just reading text chat messages, reading and writing text chat messages, or listening to others and speaking out loud – their ability to recall the specific details of the videos was diminished.

Despite the preponderance of evidence showing that viewers with a chat feature were distracted, we did not find that the levels of distraction were so high that they were ruinous to the experience. Participants consistently rated their enjoyment of the studies positively. Informal feedback on their enjoyment of the studies was highly positive, as well. Many participants with a chat feature expressed some kind preference for it, by using it (MovieLens, Cartoon, Text vs. Audio studies), using it during the videos even though they had the option to chat after the videos had ended (Cartoon study), or reporting that they preferred it to not chatting (Text vs. Audio study). Further, the self-reported ratings of distraction generally fell in the middle of the scale (in the Cartoon study, $M = 3.5$ of 7; in the Text vs. Audio study, $M = 3.6$ of 7 for participants with text chat and $M = 4.2$ of 7 for participants with voice chat). In addition, our objective measure of distraction – recall of events in the videos – showed only a small difference between participants with chat and participants without chat. In the Chat Distraction study, participants with chat got about two fewer questions correct than participants without chat; in the Text vs. Audio study, participants with chat got about 1 fewer question correct than participants without chat. Therefore, we conclude that despite the distractive effects of the chat feature, the degree to which it distracts viewers (for the videos we studied) is not sufficiently high to be detrimental to their enjoyment of the collaborative online video watching experience. Note that participants in these studies were not incentivized to remember anything about the videos, as we did not reward participants for remembering more. If memory recall were incentivized, it is possible that we would not have seen any effect of a chat feature on distraction.

Intermission periods did tend to reduce participants' feelings of distraction. This finding suggests that collaborative online video sites can help viewers manage their attention by providing structure to the experience. Such structure may be necessary

in cases where the video content requires more cognitive effort to process, such as political debates or news programs, or when viewers have a high emotional attachment to or great anticipation of watching a video. In these cases, a viewer's need to recall the information in the videos, or to immerse themselves by watching with minimal distractions, may outweigh their need or desire to socialize with other viewers.

10.3. DESIGN RECOMMENDATIONS

Many of the participants in the studies enjoyed chatting while watching videos, despite it being distracting. In addition, participants enjoyed using both text and voice chat features. These findings provide strong support for including real-time chat features in a collaborative online video watching experience. The social aspects of the experience seem to outweigh any negative consequences of distraction.

However, chatting while watching videos is not a universally appealing activity. A small number of participants in each study reported simply not being interested in the chat feature. This sentiment was best summarized by a participant in the MovieLens study who said, "I'm not interested in chatting online, especially not during watching a movie" (ML6). In addition, small percentages of participants in the Text vs. Audio study reported that they would have preferred not to chat while watching the videos (between 3-8% of the participants in the chat conditions, and 33% of the participants without chat).

The presence of participants who report not being interested in a chat feature while watching videos, even after they have experienced it, suggests that we ought to be careful in how we design collaborative online video watching experiences. Social features, such as a text chat or a commenting system, should be clearly explained to viewers and presented in a way that piques their interest, rather than in a way that overwhelms them and causes them to feel immediately distracted or overwhelmed by their presence. The presentation of these features in the user interface can also be scaffolded, such that viewers are first exposed to lower-bandwidth social features like status indicators that show what their friends are watching or chat summaries that do not require them to produce their own chat. Over time, as viewers become more comfortable with having social features while watching, they can opt to use higher-bandwidth features like real-time text or voice chat. In Chapter 11, I discuss

the use of lower-bandwidth visual summaries of chat messages. These summaries require that viewers only look at them occasionally to gain a sense of what other viewers are talking about. They were designed to avoid the overwhelming effects of having many chat messages scroll past in an active chat channel by aggregating and/or filtering the messages. Thus, these summaries could be used to introduce newcomers to the notion of watching collaboratively by showing them what other people are saying in a non-overwhelming manner.

10.4. SUMMARY AND CONCLUSIONS

- Chatting while watching provides a sociable experience for viewers and was enjoyed by groups of friends and groups of strangers in the laboratory studies.
- The videos act as a “social glue” that can bring people together online and encourage them to engage in a social experience by chatting with other viewers.
- The videos do not discourage viewers from talking about personal topics, which suggests that collaborative online video watching can be used in an online community to promote bonding between members.
- Despite the distraction inherent to multitasking between watching videos and chatting, the magnitude of the distraction is small, and viewers are willing to cope with being distracted in order to have the social experience of chatting with their friends while watching.
- Not all viewers want to chat while watching. Video player interfaces should scaffold social interaction features to help viewers manage their attention. Visual chat summaries, such as those discussed in Chapter 11, can provide this scaffolding.
- Real-time chat, embodied in either textual or auditory media, is a viable design option for online video sites that wish to promote social interactions amongst their members.

Part III: Large-Scale Collaborative Watching

Popular online video events often attract millions of simultaneous viewers. Part III discusses several challenges individuals face when participating in a large virtual audience: finding groups of people with whom to chat, maintaining an awareness of the activity of other audience members, and feeling a connection to those other members. To address these challenges, I evaluate several designs for representing large audiences and summarizing their chat. I conclude by presenting a design for a collaborative online video site that supports a large audience of viewers watching together in real-time.

11.

DESIGNING FOR LARGE VIRTUAL AUDIENCES

The introduction to this dissertation began by recalling the online video broadcast of President Obama's inauguration. This event was monumental because it signaled both a market for and a willingness of broadcasters to provide live, streaming video to an audience of 7.7 million viewers. It also signaled a growing recognition of the value of social interaction around video by integrating the broadcast with Facebook and allowing viewers to post status message updates that were shared with the entire audience (Sutter, 2009).

Television content often attracts audiences of millions as well. Table 11-1 lists several popular online video and television events and their estimated viewership. As these events become more commonplace, and as they become available online, it becomes important to consider how to design enjoyable and engaging experiences for large online audiences. In this chapter, I consider two challenges when designing collaborative online video experiences for large audiences of simultaneous viewers: how viewers find other viewers with whom to chat, and how viewers maintain awareness of the entire audience of viewers.

Table 11-1. Popular television and online video events that have attracted audiences of millions. The source of the estimated viewership is listed, along with the date the event occurred.

Event	Date	Est. viewership
CNN/YouTube Debate in S. Carolina	Jul. 23, 2007	TV: 2.6 mil (Gough, 2007)
Obama's Speech on Race	Mar. 18, 2008	YouTube: 3.8 mil (Melber, 2008)
Lost (online episode views)	Dec. 2008 (entire month)	Online: 1.4 mil (Whitney, 2009)
Inauguration of President Obama	Jan. 22, 2009	Online: 7.7 mil (Sutter, 2009) TV: 37.8 mil (Gough et al., 2009)
Super Bowl XLIII	Feb. 1, 2009	TV: 98.7 mil (Nielson)
WWDC Keynote Address 2009	Jun. 8, 2009	Ustream.TV: 20k+ (self-participation)

11.1. THE CHALLENGES OF A LARGE AUDIENCE

The studies in Part II examined the small-group experience of watching videos online and chatting with others. They found that the experience was enjoyable when viewers chatted in small groups of 2-4 people. The goals of Part III are to understand the challenges present when trying to provide a similarly enjoyable experience to a large audience of viewers, and to design features in the user interface that help people manage their social interactions with others as they watch.

It is a fundamental tenet of democracy that people are able to express their opinions openly and freely, and that ideas can be discussed and debated by all participants. However, democracy doesn't scale, and one reason for this is because it is hard to allow everyone in a large audience to freely express their opinions. Chatting online suffers from this same problem. When millions of people all talk at once, who is left to listen? Thus, chatting with all members of the audience simultaneously seems unrealistic. With an abundance of people with whom to chat and messages to read, viewers will face problems of information overload. Having too many chat options may lead to feelings of distraction and frustration, and discourage use of the chat feature. Therefore, this chapter addresses the following challenges:

- Finding viewers with whom to chat (Section 11.2), and
- Maintaining an awareness of other viewers (Section 11.3).

Before addressing these challenges, we must first understand how a chat feature will operate for a large audience. Do viewers conduct back-and-forth conversations with each other, or do viewers simply participate in an aggregated-and-filtered message stream?

11.1.1. CHATTING IN A LARGE AUDIENCE

There are several models for providing a chat feature to viewers in a large audience. Two popular models are the aggregate-and-filter model and the chat group model. This section discusses both of these models, as well as a hybrid model that enables viewers to have a social watching experience without being overloaded, and without losing awareness of the entire viewing audience.

Aggregate-and-filter. In the aggregate-and-filter model, messages from viewers are first centrally aggregated and then distributed to other viewers. To avoid information overload, these messages are filtered for each user. This model is currently used by the Facebook Live Stream Box (Siegler, 2009). Status message updates are aggregated by Facebook, filtered for each user, and distributed to viewers in a periodically-updating visualization¹⁷. Figure 11-1 shows an example of this visualization. Although the technical details of how this widget chooses messages to display have not been made public, the widget does seem to prioritize messages that originate from within a viewer's social network. In addition, this widget was designed to support a viewership of millions, although the developers indicate that in this case, not all viewers will see all of the messages that are posted.



Figure 11-1. The Facebook Live Stream Box. This widget can be coupled with any live video stream online to create a branded, collaborative video experience for viewers with Facebook accounts. Names and photos have been blurred to preserve anonymity.

¹⁷ My observations of this feature during several live events were that two to four new status messages were added to the list every seven seconds.

The Live Stream Box can handle a very substantial load, supporting millions of simultaneous users. The information may move fast, so users will not necessarily see everything that gets posted and cannot page through all the history. (from the Facebook Developer Wiki, http://wiki.developers.facebook.com/index.php/Live_Stream_Box)

Because viewers may not see all of the messages sent, back-and-forth conversation may be difficult or impossible. When messages are filtered out, it is unclear to message senders who will see their message. Further, for viewers who reply to a message, it is unclear if the original sender will see the reply.

Chat groups. An alternative to the aggregation-and-filter model is to explicitly group viewers into smaller chat groups with explicit group membership (e.g., IRC-style). In this case, viewers are explicitly in one chat group and not in others. The chat group model preserves the ability for viewers to conduct back-and-forth conversations, and when these groups are kept small, can provide an experience similar to those studied in Part II.

Segmenting a large population into many smaller groups will reduce the information-processing burden on reading too many chat messages. It may also incentivize viewers who would otherwise lurk in a larger chat group. Visibility within a group is one way to elicit contributions from group members (Karau & Williams, 1993; Oliver & Marwell, 1988). By segmenting the population into smaller groups, people may feel more inclined to contribute to the conversation because they are more visible to the other viewers in the group and their contributions are more likely to be seen. However, by excluding viewers from other groups, they are disconnected from the audience as a whole.

Another consideration for the chat group model is the limit on the size of the groups. Researchers have examined the effects of group size on the ability for members to conduct back-and-forth conversations. Sacks, Schegloff, and Jefferson (1974) report that in face-to-face conversations, small groups of three to four people are able to maintain a conversation on a single topic. In larger groups, it is common for 'schisms' in topics to occur. These schisms fragment the conversation into two or more threads. Participants may simultaneously contribute to one or more of these threads, which may continue apart, end, or re-join into a single topic. In an unstructured conversation, as more people join the conversation, the likelihood that

conversational schisms occur increases. In addition, as the number of schisms increases, so does the likelihood that their topics drift farther apart. O'Neill and Martin (2003) report similar effects for computer-mediated text chat channels with 6 to 11 participants. Therefore, there are limits to the number of people who can participate in a chat; once those limits are crossed, separate conversational threads emerge among subgroups of participants.

Jones et al. (2008) performed a quantitative study of participation in IRC text chat channels to determine this 'carrying capacity' for text chat. They found several interesting limits. First, they found that IRC channels were limited to about 300 concurrent users (including lurkers), with fewer than 40 active users (non-lurkers). Second, they found a limit to the amount of activity that could be sustained in each channel: about 600 messages per 20-minute interval. This rate corresponds to about 30 messages per minute, or 1 message every 2 seconds; increases in message traffic beyond this limit may result in information overload. Finally, they found a negative effect of group size on the distribution of participation among group members. As the number of users in a channel increased from 14 to 150, the ratio of chatters to channel members decreased from 1.0 (everyone chatted) to 0.2 (only 20% of users chatted). In other words, in a group of 150, only 30 users are inclined to use the chat feature. Therefore, the types of chat groups that seem most capable of providing viewers with a social experience without overwhelming them are those that are small (< 40 active users) and active (one message every few seconds).

Hybrid model. The aggregation-and-filter and chat group models are not mutually exclusive. A hybrid model can preserve the best features from these models – back-and-forth conversation and large-audience awareness. To enable back-and-forth conversation, viewers chat inside of explicit chat groups, similar to the groups discussed in Part II. To enable awareness of the audience as a whole, visual summaries of chat and social proxies that summarize audience activity are employed. To aid viewers in finding interesting groups to join, a chat group recommender system is used. I discuss these features in the rest of this chapter, and conclude with results from a user study that describes peoples' impressions of a prototype user interface that incorporates the main features employed by the hybrid model.

11.2. FINDING CHAT GROUPS

In an audience of millions, the number of chat groups required to provide a small-group setting for viewers is in the thousands. For example, using the limit of 300 users found by Jones et al. (2008), the audience of 7.7 million viewers watching President Obama's inauguration would have been distributed among over 25,000 chat groups. Using a limit of 40 users, the number of groups required is over 190,000. How does a viewer choose which group to join when there are so many available options?

There are several approaches, representing a design space in how much control the user has in selecting a group. On one extreme, group selection can be fully interactive by having each user view the entire list of groups and select the one they want to join. On the other extreme, group selection can be fully automatic by having the system assign each user to a group based on some pre-defined metric. Both of these methods create significant problems for users.

In the fully interactive case, users will face a problem of information overload when selecting a group, as they must select from hundreds to thousands of options. Prior work I have done showed that users are particularly bad at picking groups of interest when there are hundreds of options, even when those users have a specific information-seeking goal and group names are descriptive of the information they are trying to find (Weisz, Erickson, & Kellogg, 2006). In that study, users simply gravitated to one well-known group (called "everyone"), producing an abundance of participation in that group, and a lack of participation in other, more topically relevant, groups. We expect the same behavior in an audience of millions of online video viewers. In this situation, viewers do not have the same information-seeking goal; rather, their goal is to find an enjoyable social experience. Evaluating which groups are likely to provide this experience can be difficult, if not impossible, when viewers are presented with too many options. Thus, we can expect viewers to congregate in the most active groups, producing a small set of unsustainably large groups.

A group assignment strategy that load-balances users into chat groups can ensure that group size does not extend beyond its natural limits. However, when users are automatically assigned to groups, they may find their assignments unsatisfactory for a variety of reasons: they may not like the people they are grouped with (e.g., some

Democrats may be upset if grouped with Republicans), they may not find the chat interesting, or they may want to chat with people they already know (e.g., their friends or relatives). Therefore, a system that automatically assigns users to chat groups should take into account users' individual preferences for the type of group for which they are looking.

A hybrid scheme that provides recommendations for groups based on each users' preferences, but leaves the final choice of group to the user can overcome the limitations of the fully-interactive and fully-automatic schemes. In this scheme, the user remains in control of their group assignment. Further, the system can still perform load-balancing by filtering out recommendations for groups that have too many users, or by promoting groups that have too few users.

How should such a group recommender system be built? What are the different dimensions of chat groups that can be used as a basis for making recommendations? Which of these dimensions are most important to people? How should these dimensions be weighted by a recommender system to make a set of recommendations? Table 11-2 provides a list of many of the chat group dimensions that can be used as a basis for making recommendations. Note that some of these dimensions, such as group size and chat activity, are common to all types of Internet chat groups and not just those focused on watching videos. In the case of social networks, we assume that social network data are available in the chat system.

Table 11-2. Dimensions of chat groups that can be used as a basis for recommending those groups to users.

Dimension	Description
Group size	Number of members in the chat group
Chat activity	Average number of messages per minute
Conversational topics	Recommendations can be made for groups in which people are talking about similar or dissimilar topics as the user or the user's current group
Geography	Recommendations can be made for groups based on the locales of their members
Video synchronization	Video synchronization deals with the relative difference between viewers in the position of their video playback; recommenders can use this feature to ensure that chat groups consist only of viewers who are relatively synchronized with each other (i.e., seeing the same part of the video at the same time)
Social network	Recommendations can be made based on whether a user's friends or friends of friends are in the group



11.2.1. EVALUATION






















I conducted a survey to understand the types of chat groups in which people would be most interested as well as the relative importance of each of the chat group features in making group recommendations. Although self-reported preferences are not as accurate as real-world behaviors, they do provide guidance for which features should be explored when actually creating a chat group recommender system.

The survey consisted of two parts. First, respondents were asked about their interest in particular types of chat groups for each dimension listed in Table 11-2. For example, for the group size dimension, respondents were asked which type of group they would prefer: a small, medium, or large group (actual sizes were given as listed in Table 11-3). Except for the social network dimension, respondents could choose as many preferences as they liked; for social networks, the preference choices were mutually exclusive and respondents could only choose one. Next, respondents were asked to weight each dimension based on how a chat group recommender system should use that dimension in making a chat group recommendation. To make the weightings, respondents were given 60 points to distribute among the six dimensions.

Respondents. This survey was administered at the end of the user study presented in Section 11.5. Ten people responded to the survey, half of whom were female. The mean age of respondents was 22.2 (SD = 2.8) years, and respondents estimated they had a mean of 549 (SD = 474) friends on Facebook. Respondents were instructed in how to make the weightings and reminded that they could choose multiple types of groups within each dimension (e.g., for group size, they could pick both small and medium groups if they liked). Table 11-3 reports the preferences reported by respondents, in the order presented on the survey.

Table 11-3. Self-reported preferences for different features of chat groups. Weights are the sum of the weights assigned to each category across all respondents. Respondents could express interest for multiple types of groups in each category except for (*) social network. () This answer choice was not present on the survey and was written in by two respondents.**

Type of group	# interested	Σ Weight	Rank
Group size		97	4
A small group (2-8 people)	 6		
A medium group (9-20 people)	 5		

Type of group	# interested	Σ Weight	Rank
A large group (> 20 people)	 3		
Chat activity		116	2
A quiet group (no chat)	 0		
A moderately active group (a few messages every few minutes)	 2		
An active group (a few messages every minute)	 8		
A highly active group (a few messages every second)	 2		
Conversational topics		84.5	5
A group that talks about topics you have previously talked about	 6		
A group that talks about topics you have not previously talked about	 5		
A group that talks about a specific topic you have in mind	 7		
Geography		62.5	6
A group with people in Pittsburgh	 4		
A group with people in Pennsylvania	 0		
A group with people in the US	 7		
A group with people from another country	 3		
Video synchronization		100	3
People in the group must be almost fully in-sync	 6		
People in the group can be ahead or behind by 1-3 seconds	 3		
People in the group can be ahead or behind by more than 3 but less than 10 seconds	 2		
People in the group can be ahead or behind by 10 seconds or more	 1		
Social network*		140	1
I prefer not to chat at all	 0		
I prefer to chat only with my friends	 5		
I will chat with strangers as long as I have at least 1 friend in the group	 3		
I prefer chat with both friends and strangers**	 2		
I prefer to chat only with strangers	 0		

Group preferences and feature importance. Overall, respondents expressed a strong interest in watching videos with their friends. The social network was rated as the most important feature when finding a chat group, and all respondents expressed a desire to watch with their friends, either alone (50%), or in the company of strangers (50%). Thus, a chat group recommender system should strive to route friends to the same groups, although the willingness of respondents to chat

with strangers in the presence of friends indicates that chat groups can also be formed by combining small groups of friends from different social networks (e.g., creating a chat group by combining two friends from one network, three friends from another, and two friends from a third). In this way, interactions between strangers are encouraged.

The activity level in chat is another important feature to consider when recommending chat groups. Respondents preferred active groups, with a specific preference for groups having a few messages arrive every minute; this preference is also reflected in the fact that no respondents reported not wanting to chat in the social network question. Interestingly, no respondents reported preferring quiet groups with no messages, although this observation may be an effect from the small sample size. However, the relative ranking of chat activity indicates that a chat group recommender system should ensure that each group has active members, and avoid creating groups composed solely of lurkers.

The next important feature for recommending chat groups was video synchronization. Respondents strongly felt that they needed to be in a chat group with other viewers who were synchronized with them. Only 3 respondents felt that a skew of 3-10 seconds or a skew of more than 10 seconds in video playback was acceptable.

Group size was the next important feature to consider when recommending chat groups, although it was not as important as the activity in chat. Overall, respondents reported preferring small-to-medium sized groups having roughly 2-20 members. Only 3 respondents preferred large groups with more than 20 members. This finding suggests an even tighter upper bound on group size than that suggested by Jones et al. (2008); empirically, the bound is about 40 active chatters, whereas preferentially, the bound seems to be at least half that. Note that the bound reported by participants may have depended on the nature of the event. In this study, participants were primed to think about political events because a screenshot from the inauguration broadcast was used. Participants may have reported a different preference in the case of other types of events, such as watching a YouTube video or a highly-anticipated movie. Despite this event bias, we still recognize the upper bound of about 40 active chatters found in the study by Jones et al. (2008).

Respondents were mixed in terms of their preferences for conversational topics, although generally, conversational topics were not rated as being important. Of the 6 respondents interested in chatting about topics previously chatted about (i.e., similar topics), 4 were also interested in chatting about topics not previously chatted about (i.e., different topics). More important was the ability to find groups chatting about a specific topics respondents had in mind, suggesting the need for a search feature that allows users to find groups based on their topics of chat.

The questions about geography were designed to understand the scale at which people were interested in meeting others – would they want to meet people in their same city (Pittsburgh), their same state (Pennsylvania), their same country (U.S.), or would they want to meet people in different countries? However, while running the study, it was discovered that these questions were confounded with the respondent's identity with Pittsburgh, Pennsylvania, and the city in which they grew up. Most respondents were students temporarily living in Pittsburgh; only one was a native Pittsburgher. In addition, several students were from other countries, and their interest in chatting with others from other countries likely reflected an interest to chat with their friends and family living in those countries. However, as geography was ranked last in terms of importance, the only reliable conclusion seems to be that respondents cared about geography inasmuch as it related to where their friends and family live.

11.3. MAINTAINING AWARENESS OF THE AUDIENCE

One consequence of segmenting viewers in a large audience into many different chat groups is that each viewers' awareness of the rest of the audience is diminished. If no indication of the size or presence of the audience is given, viewers in one chat group may be wholly unaware that there are other viewers in other chat groups. However, even with an indication of audience size, viewers in a chat group may be completely unaware of the composition and activity of that audience – who are these people and what are they talking about? This lack of information may cause audience members to feel disconnected from the group as a whole. Awareness tools that summarize an audience and its activity can give individual members stronger feelings of presence and connectedness to the entire group (Erickson & Kellogg, 2000; Weisz, Erickson, & Kellogg, 2006; Ren, Kraut, & Kiesler, 2007).

11.3.1. VISUALIZATIONS OF USERS AND ACTIVITY

Many visualizations have been created to summarize large quantities of data, help people locate information, and provide awareness of other people and their activities online (e.g., Erickson et al., 1999; Kellogg et al., 2006; Halverson et al., 2001). These visualizations support two types of awareness: awareness of individual users or groups in the system and their status, activities, and/or interactions with each other; and awareness of the contents of their interactions (i.e., their chat or status messages).

Individuals and groups. Social proxies are minimalist visual representations of individual users and their activities in a system. They are designed to improve coordination and accountability in online communication and collaboration spaces by increasing the visibility of each user’s activity to other users. Such systems are described as being “socially translucent” because they make users’ social behaviors – their interactions with the system and with each other – visible to all users in the system (Erickson & Kellogg, 2000). Social proxies have been designed for several real-time communication and collaboration systems, detailed in Table 11-4.

Table 11-4. Visualizations that provide awareness of users and their status, activities, and/or interactions with each other. The design elements used for representing users, their status, and their activities in the system are given.

System	Users	Status	Activity	References
Babble & Loops <i>online discussions</i>	Colored circles	In/out of the discussion group: placement inside or outside of the “cookie”	Proximity to the center of the cookie shows recency of activity	(Erickson et al., 1999; Erickson et al., 2006)
Rendezvous <i>conference calls</i>	Semicircles (with names)	On/off of the conference call: placement at the virtual table vs. on the waiting list	Orange highlight for people currently speaking on the conference call	(Kellogg et al., 2006)
Task Proxy <i>workgroups</i>	Colored hexagons	Task state: color represents no task state, task in progress, or task completed	None	(Erickson et al., 2004)

System	Users	Status	Activity	References
Activity Map <i>large-scale discussions</i>	Users not represented individually	None	Users' forum posting activity aggregated into circles plotted on a world map	(Halverson et al., 2001)
Ellis Auditorium <i>online lectures</i>	White dots	Group membership: white dots are placed on top of blue circles representing chat groups	None	(Olguin & Kruper, 2004)

Babble, and its successor Loops, both contain a social proxy that represents users and their activities on a message board (Erickson et al., 1999; Erickson et al., 2006). This proxy is known informally as “the cookie” because of the large circle into and around which users are placed. Individual users are represented by small circles, and their distance to the center of the cookie indicates how recently they posted a message to the discussion; closer proximity indicates more recent activity. Users who are browsing other topic threads in the system are displayed outside of the cookie. The cookie visual was designed to increase awareness of group activity and facilitate communication among members.

Rendezvous is a system that helps users schedule and participate in conference calls (Erickson et al., 2006). Because audio channels lack cues about peoples' status – who is connected, who is speaking, and who is muted – Rendezvous uses a social proxy to display information about the people on the call. Attendees are represented by gray semicircles seated around a virtual table. When someone speaks, their semicircle grows to a full circle and changes its color to orange. Attendees who are muted have their names grayed out. Finally, attendees who have not yet joined the call are displayed in a separate waiting area. In a user study of this system, Ding et al. (2007) found that users found the speaking indicators helpful, especially for identifying people when their voice was unfamiliar. They also found that some users took screenshots of the proxy to keep track of meeting attendance.

The Task Proxy represents team members in a workplace and displays their progress to completion on a group task (Erickson et al., 2004). It was designed to provide managers with an overview of the status of their projects, as well as increase users' motivation and feelings of accountability by making visible their state and the state of their co-workers. In this proxy, users are represented by

hexagons that are joined together to form a ‘honeycomb’ structure, representing the team. Hexagons are colored according to task completion state: no state entered, task in progress, or task completed. In a user study of this proxy, Erickson et al. (2004) found that many participants found the concept useful and generally gave positive ratings to the proxy’s ease of use and value.

Another visualization that represents users inside of groups is called “Cosmo,” used by Ellis Auditorium (Olguin & Kruper, 2004). Ellis Auditorium is a system that supports watching online lectures and chatting with other viewers. The Cosmo visualization uses circles to represent chat groups, and white dots inside those circles to represent individual users. The visualization reconfigures itself depending on the number of available chat groups. As with the social proxy visualizations, Cosmo represents each individual user, and was designed to support an audience of dozens of users distributed among a handful of chat groups.

The Activity Map was designed for use in the WorldJam event at IBM in 2001 (Halverson et al., 2001). WorldJam was a 72-hour online event during which all IBM employees could connect with each other by asking and answering questions, and participating in discussions. The Activity Map visualization was used to help employees navigate through the discussion fora, and to provide mutual awareness of the activities of other employees throughout the world. It did this by highlighting currently active forums and by displaying aggregate participation data on a world map.

Many of the virtual group visualizations discussed were designed for small groups of users. These include the Babble cookie, the Rendezvous proxy, and the Cosmo visualization. In each of these designs, individual users are represented concretely in the visualization – such as by a circle, hexagon, or dot – and information about their status and/or activity is presented using color and/or position. Because they render each user individually in the visualization, they are unable to directly scale to audiences of millions; such a scale would require millions of pixels in screen real estate, which is contradictory to the minimalist design aesthetic of a social proxy. In Section 11.4, I present the design of a new social proxy capable of representing the activity of an audience of millions. This social proxy follows similar principles as the Activity Map by representing aggregate groups of users and displaying information about their activity level in the system.

Content of interactions. In addition to explicitly representing users and their state in a system, a second type of visualization can be used to summarize the contents of their interactions, such as their chat and/or status messages. Several common visual summaries that have been applied to chat and/or status messages are described in Table 11-5.

Table 11-5. Visualizations that summarize the contents of users' interactions, such as their chat or status messages.

Visual summary	Information	Examples
Tag cloud	Popular keywords; often used as a navigational aid on web sites	delicious.com/tag , flickr.com/explore , technorati.com/tag
Scrolling list	Individual messages presented in a list format; new content added periodically over time	Facebook Live Stream Box, Twistori (Figure 11-3)
Spatial / interactive	Individual messages presented spatially; requires interaction to read messages	We Feel Fine (Figure 11-4)

Perhaps the most ubiquitous visual summary of information is the tag cloud. Tags are short snippets of text that are descriptive of an item. For example, on a photo-sharing site, a user may tag a photo with “sunrise” and “hawaii” to describe a photo she took of a sunrise in Hawaii. Tag clouds are used to summarize the relative popularity of the tags in the system. An example tag cloud is shown in Figure 11-2. Often, tags are formatted to display their relative importance, such as by changing their size, color, or position. Tags and tag clouds have been used in a wide variety of contexts, including summarizing the topics of a blog¹⁸, displaying the most popular tags applied to photos¹⁹, and web site bookmarks²⁰. They have also been used to show the relative popularity of classes in the Java API (Stylos, Myers, & Yang, 2009). Thus far, tag clouds have not been applied to the contents of chat messages. An evaluation of the use of tag clouds for summarizing chat is discussed in Section 11.5.

Another method for summarizing information over time, such as chat messages and status message updates, uses a scrolling list format. As discussed earlier in this chapter, the Facebook Live Stream Box (Figure 11-1) uses this format: status messages are selected according to some metric and presented in their full,

¹⁸ Technorati. <http://technorati.com/tag>

¹⁹ Flickr. <http://flickr.com/explore>

²⁰ Delicious. <http://delicious.com/tag>

scrolling list or tag cloud: users must actively click on shapes to read the messages they represent.

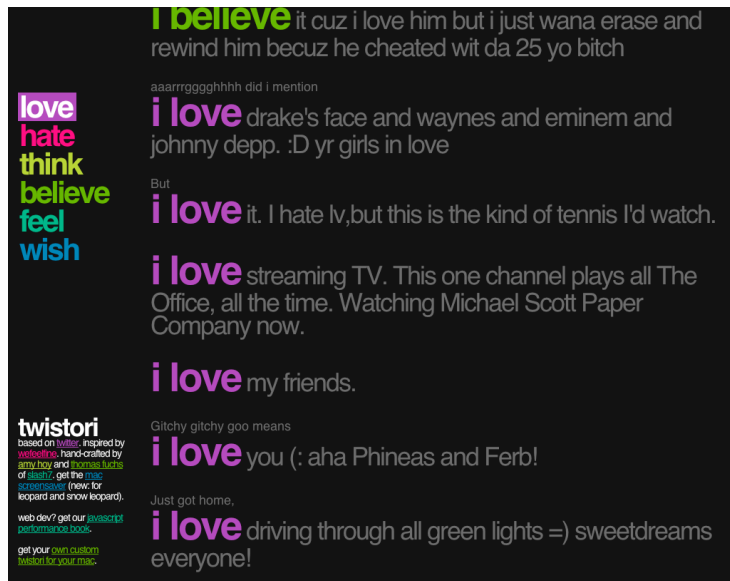


Figure 11-3. Twistori visualization of messages on Twitter. Messages are selected on the basis of containing phrases such as “I feel” or “I think.” New messages are added at a rate of about one per second.



Figure 11-4. We Feel Fine visualization of feelings and sentiments made in blog posts. Individual messages are represented with small icons that swarm around the interface. Messages can be viewed by clicking on them. Color represents the feeling described by the message.

Because of the high degree of interactivity required, the spatial layout visualizations may not be appropriate when used in a collaborative online video context; if simply reading chat while watching videos is distracting (Chapter 7), then any visualization requiring even more attention from the viewer will add to their level of distraction. Therefore, in evaluating visualizations that summarize the chat messages of a large, virtual audience, I consider passive visualizations that do not require user interaction, such as the tag cloud and scrolling list. Despite the risk of being additionally distracting, participants in the user study described in Section 11.5 reported interest in interactive visualizations. Thus, further study of the distraction of interactive visualizations is required. I discuss this point further in Section 17.4.

11.3.2. DESIGN FOR A LARGE, VIRTUAL AUDIENCE

Awareness tools such as social proxies, tag clouds, and scrolling lists of chat messages can be applied to the collaborative online video context to provide a fragmented audience with an awareness of other viewers. Both types of awareness are important for viewers watching a live event in a large audience: an awareness of the actual audience members (representation), and an awareness of their topics of conversation (summarization). Visual summaries such as tag clouds and scrolling lists of chat messages can be directly applied to the large-scale collaborative online video experience, as the individual chat messages typed by viewers are the only required data source. Representing an audience with a social proxy is more difficult, as current proxies cannot be directly applied; they do not scale to millions of users. In the next section, I describe the design of a social proxy for large audiences. Following this design, I report the results of a study that evaluates this proxy, as well as the tag cloud and scrolling list chat summaries.

11.4. A SOCIAL PROXY FOR LARGE AUDIENCES

The social proxy concept has been shown to work well for representing individual users and their interactions in a group (Erickson et al., 1999; Erickson et al., 2004; Erickson et al., 2006; Kellogg et al., 2006). However, the social proxies presented earlier generally suffer from the same limitation with regard to large audiences: they represent users as individual graphic elements, and hence cannot scale to an audience of millions. In this section, I present the design of a novel social proxy that represents audiences of this magnitude.

11.4.1. THE RADAR CONCEPT

The general concept of the audience proxy is shown in Figure 11-5. It follows a “radar” concept, showing the composition of the audience from the perspective of the individual user. It was designed to display a large, virtual audience compactly, such that it could be embedded as only one small portion of a larger user interface.

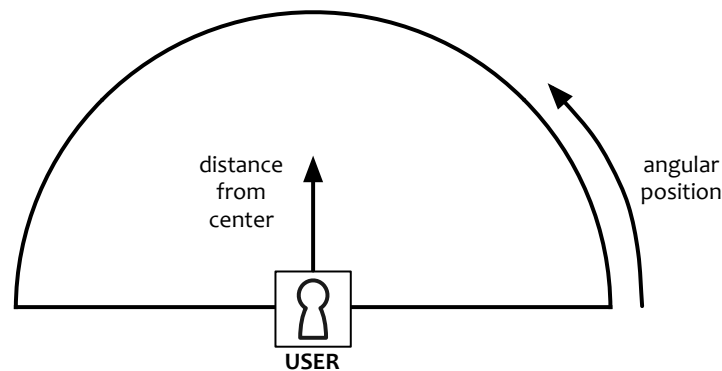


Figure 11-5. General concept for the large audience proxy. The audience is presented from the perspective of an individual user. Two dimensions are used to place audience members: angular position and distance from the center.

The radar concept has two degrees of freedom for placing representations of other audience members: angular position ($0^\circ - 180^\circ$) and distance from the center. These freedoms in positioning can be used to convey information about audience members. For example, Babble used distance from the center to convey recency of activity. The same capabilities exist in this concept as well. Note that because angular position and distance are measured on a continuous scale, continuous attributes can be mapped onto them.

In addition to representing continuous attributes, the proxy can represent ordinal or nominal categories by discretizing the attribute space. Examples of the discretization of both angular position and distance from the center are shown in Figure 11-6. Discretization can either be into equal-sized categories (Figure 11-6a) or into categories with sizes proportional to the number of people who fit into each category (Figure 11-6b).

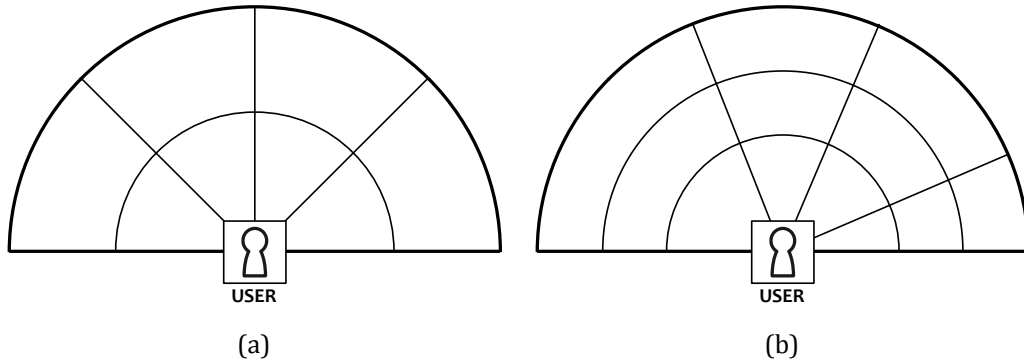


Figure 11-6. Discretization of angular position and distance to center for nominal or ordinal attributes. (a) Four angular categories and two distance-based categories of equal size. (b) Four angular categories and three distance-based categories with areas proportional to the number of audience members in each category.

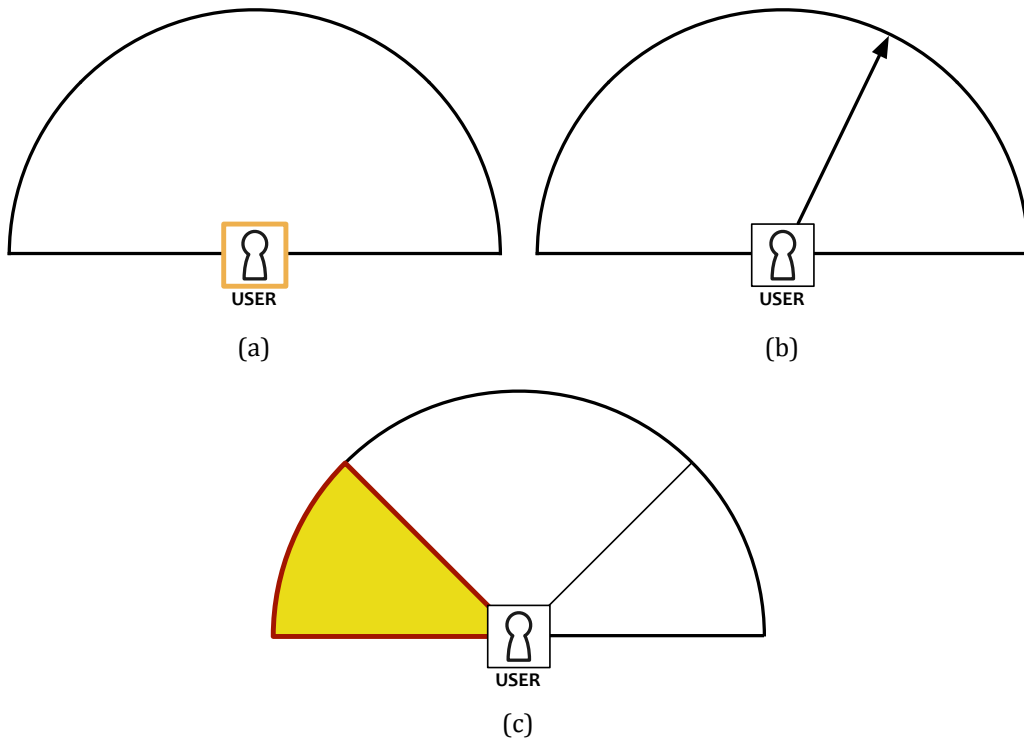


Figure 11-7. Using color to encode attributes of the user. (a) A color highlight around the user's icon can represent an attribute encoded by distance to the center. (b) A line can be used to represent a continuous attribute represented by angular position. (c) A border, edge, or background highlight can be used to represent an ordinal or nominal attribute of the user.

As for representing audience members, in cases of small audiences, icons or graphics are used to represent all individual audience members. In cases of large audiences, icons or graphics are used to represent clusters of audience members. These clusters are comprised of members who are grouped together on the basis of similarity in the metric spaces. As this proxy was designed for large audiences, I focus only on this case; however, it can be adopted for small audiences as well.

One last detail is how to represent the attributes of the individual user. Because the proxy places the individual user in a fixed position, at the origin point of the radar, the user's attributes cannot be represented positionally. Thus, other characteristics such as color and highlighting are used. Figure 11-7 shows several designs for encoding the user's attributes.

As for representing audience members, in cases of small audiences, icons or graphics are used to represent all individual audience members. In cases of large audiences, icons or graphics are used to represent clusters of audience members. These clusters are comprised of members who are grouped together on the basis of similarity in the metric spaces. As this proxy was designed for large audiences, I focus only on this case; however, it can be adopted for small audiences as well.

One last detail is how to represent the attributes of the individual user. Because the proxy places the individual user in a fixed position, at the origin point of the radar, the user's attributes cannot be represented positionally. Thus, other characteristics such as color and highlighting are used. Figure 11-7 shows several designs for encoding the user's attributes.

This audience proxy breaks a common design convention for social proxies, that all users share the same representation of the audience. The reason for this departure from convention is because the user is the focal point of the representation and other audience members are displayed according to their relationship to the user. Shared representations are desirable because they support mutuality and accountability; since everyone can see everyone else's state, everyone knows that everyone else knows their current state. For example, an under-contributing team member may be motivated to contribute more when he realizes that others can see that he is under-contributing.

In the case of collaborative online video watching, the audience members are not in a situation in which they are being held accountable for anything. Thus, it is less

important to provide a consistent, global view of the audience. Collaborative watching is meant to be fun and social, and the goal of the audience proxy is to provide viewers with a feeling of connection to the rest of the audience. Individualized audience representations can provide this connection. Having an individualized representation also enables users to interactively explore the composition of audience members and discover how other audience members relate to them. These interactions are discussed further in Section 11.5.3.

11.4.2. PROTOTYPE FOR A LARGE, VIRTUAL AUDIENCE OF VIDEO WATCHERS

An example audience proxy for video watchers in a large, virtual audience is shown in Figure 11-8. In this proxy, the attributes that are encoded are the level of activity in chat (angular position) and the relationship of the audience members to the user (distance to center). These attributes were chosen because they seem to represent interesting aspects of the social viewing experience. “Chattiness” gives viewers an idea of how many people are using the chat feature versus how many are simply watching the video. Relationship is encoded by positioning the user’s friends closer to the viewer and positioning unknown strangers as anonymous clusters of users farther away. In addition, pictures of the user and the user’s friends are displayed, instead of generic icons or shapes, to give the user a greater sense of connection with their friends. Finally, a distinction is made between friends inside of the current chat group (inner semicircle) and friends in other chat groups (outer periphery of the inner semicircle). This distinction was motivated by the design of Babble: users browsing the same message thread are “in” and users browsing other message threads are “out”.

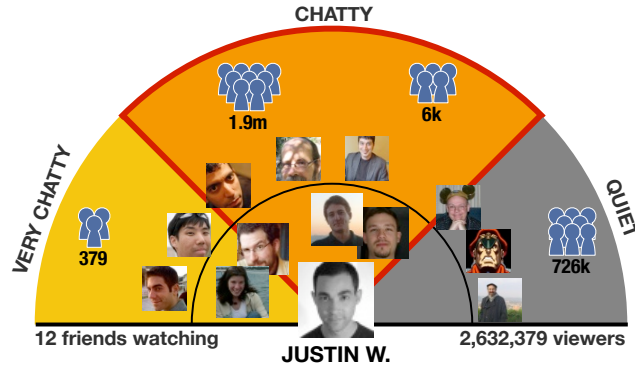


Figure 11-8. Prototype of the audience proxy for watching online video in a large audience. Chattiness is a continuous measure encoded in angular position. Relationship is a discretized measure (friend vs. stranger) encoded in distance to center. The viewer and his friends are displayed using their pictures to convey a greater degree of presence and connection; strangers are displayed using generic icons. Friends are also placed “in” the current group or “out” in other chat groups.

11.5. EVALUATION OF AUDIENCE REPRESENTATION AND CHAT SUMMARIZATION

To evaluate the ideas of representing a large, virtual audience with a social proxy, and summarizing chat messages from that audience, I developed a prototype user interface that puts these ideas together in one space (Figure 11-9). This interface can be configured to display two types of audience representations: the full audience proxy (Figure 11-8), or a simpler audience representation that just displays a counter for the number of audience members. The latter audience representation is a ‘hello world’ representation of an audience as it only gives a minimal amount of information about the audience. This representation is important to include as it is used in online spaces such as Justin.TV and the 1 vs. 100 video game on Xbox Live²³.

²³ The 1 vs. 100 game is a massively multiplayer online trivia game. It commonly attracts thousands of simultaneous players. Although it displays a virtual studio audience consisting of avatars sitting in seats, this representation only shows a small portion of the entire audience. Currently, the only representation of the entire audience is a counter of the number of players.

The interface also displays two types of visual chat summaries: a tag cloud and a scrolling list of chat messages. As this is a prototype interface, the visual chat summaries are static representations and do not update over time as they would in a real, deployed interface.

The prototype interface was implemented in Adobe Flex. To customize the interface for each participant, the Facebook Connect library for ActionScript²⁴ was used to load each participants' name, profile picture, and pictures of friends. Figure 11-9 shows screenshots of both configurations of the prototype interface.

11.5.1. PARTICIPANTS & METHOD

The Large Audience study was a think-aloud user study of the prototype interfaces. It was run to understand whether the information they conveyed about the audience and their chat was clear and understandable, and to get a sense of what people liked and disliked about the different visualization features.

Ten participants were recruited through word of mouth and the CBDR web site to participate in this user study. Five participants were female, and the mean age of participants was 22.2 (SD = 2.8) years. Participation lasted about 30 minutes, and participants were rewarded with cookies for their feedback.

Participants were first explained the situation: they were watching a live, streaming online video in a large audience, and the prototype interfaces included features that would show them information about who else was watching with them. They were asked to give their feedback on what they thought about these features – if they were useful or interesting – rather than if the specific presentation was not to their liking (i.e., if they did not like the colors). Participants were instructed to follow a think-aloud protocol when shown each interface by verbally expressing their thoughts as they examined and interpreted the interface. Handwritten notes were taken during the think-aloud portions.

Participants were first shown the interface in Figure 11-9a and the think-aloud protocol was described. The interface was presented without explanation. Participants were left to discover the concepts listed below on their own.

²⁴ Facebook ActionScript API. <http://code.google.com/p/facebook-actionscript-api/>

- The chat group contains a mixture of friends and strangers.
- The tag cloud displays the most popular words used across all chat groups in the system.
- The scrolling list displays chat messages from other audience members.
- There are over 4 million people watching the video at the same time, and 28 of them are the participant's friends.

When participants expressed confusion or uncertainty, explanations were provided. Misunderstandings were discussed in detail as they provided insight into alternative designs not originally intended. Little guidance was provided during the think-aloud process to try and elicit participants' thoughts about how the interface *should* operate by asking them, rather than by *telling* them, the design intentions. Using the Socratic method in this way, I was able to explore alternatives in the summarization features that we had not yet considered. For example, LA3 was uncertain where the words in the tag cloud came from, which prompted me to ask, "where do you think the words come from?" This question was followed by a discussion of possible sources for the words in the cloud. LA3 felt that the words in the tag cloud should have come from the video instead of the chat. This possibility was never considered during the design phase.

After the think-aloud portion was finished, participants were asked to react to the size of the audience: 4,685,385 people. They were asked if this size was "small," "medium," "large," or "gigantic," and they were asked to provide numbers for each of these sizes (e.g., "given that 4.6 million is 'large', what is gigantic?").

Next, participants were asked questions about whether the interface provided them with a "good sense of who else is watching with you," a "good sense of what the other people in the audience are talking about," and what in the interface gave them those senses.

After these questions, participants were shown the interface in Figure 11-9b and the think-aloud portion and "good sense" questions were repeated. Participants were also asked which features they generally liked and disliked, whether they preferred the tag cloud or the scrolling list chat summary, about how many Facebook friends they had, and whether they could think of any changes to the features to make them more useful or interesting.

After both interfaces had been shown and discussed, participants were asked to complete the survey described earlier in Section 11.2.1. This survey asked participants what kinds of chat groups in which they would be most interested, assuming the group they were shown in the prototype interface become uninteresting or boring to them.

11.5.2. RESULTS

Overall impressions. Overall, participants liked the idea of chatting with their friends while watching a live video. Five participants specifically mentioned they liked the ideas of seeing who else was watching, seeing their friends watching with them, and seeing what others were saying. Feelings about strangers were mixed. These feelings were best captured by LA5, LA8, and LA10, who said:

“I think it’s kind of cool to see what other people are saying besides my friends.” (LA5)

“It’s probably actually easier [to talk with strangers], they don’t have a sense of who you are so they’re not judging you at all” (LA8)

“Talking with strangers could get intense. People are stupid and could start a Facebook battle” (LA10)

Interpretation of audience size. Participants were asked to react to an audience size of about 4.6 million – did they feel that this was a small, medium, large, or gigantic audience? Interestingly, reactions were mixed. Four participants felt this size was gigantic, three felt it was large, two felt it was medium, and one felt it was small.

This question was followed up by asking participants how big (or small) the audience had to be to be classified as each of the other sizes. A summary of these audience sizes is shown in Table 11-6. Participants reported a wide range of values for each of the different audience sizes. Although the specific labels are not important, this exercise shows that participants’ perceptions of just how many viewers need to be “out there” for them to register and connect with the magnitude of audience’s size was varied. As said by LA3 in response to the 4.6 million simulated people in the interface, “there are *so many* people watching” (emphasis mine).

Table 11-6. Reported cutoffs for different audience sizes. (*) This participant referenced the population of China when thinking about what constituted a “large” audience.

Size	Cutoffs	Range
Small	100k, <50k, 1k-100k, 1k, <500k, <100, 100-200, ~1000, 100k	Hundreds to hundreds of thousands
Medium	50k-1m, 500k, 5k, 1.5m, 40m, 1k, 5k, 2m	Thousands to low millions
Large	12m, 1m, 100k, 100m, 10k, 85k, 1b*	Tens of thousands to one billion*
Gigantic	20m, 100m, 10m, 200m, >40m	Tens of millions to hundreds of millions

Audience proxy. In general, participants reported that the audience proxy provided them with a good sense of who else was watching the video. In the words of LA3 and LA4,

“This is quite informative” (LA3)

“This gives me an idea of all the audience and how active people are in talking about the video” (LA4)

Of the ten participants, only three felt that the interface with the audience proxy did not give them a good sense of the audience. However, in the case of two participants, these feelings were confounded with other parts of the interface. For example, LA9 simply preferred just knowing how many people were watching, LA5 wasn’t interested in seeing strangers in the interface (in the scrolling list), and LA1 liked the chatty/not chatty distinction but felt that his sense of the audience was only about everyone’s interest in political content.

The audience proxy was not immediately accessible to all participants. Two participants, LA8 and LA10, found it especially difficult to interpret upon first glance.

“I’m not sure of the setup of the audience members” (LA8)

“I don’t really understand what this means” (LA10)

With time and talking, these participants came to understand the proxy and the information it was telling them; thus, the proxy’s novel approach to representing an audience requires effort on part of the user to learn.

Participants made specific comments on three aspects of the proxy interface: the use of Facebook profile pictures to represent friends, the metric of “chattiness” for segmenting the audience, and the inner/outer rings of friends.

Facebook profile pictures were a problem for three participants because the pictures were a bit small and hard to see (LA8) and because it was hard to recognize friends from their profile pictures when their picture wasn’t a headshot or when they had recently updated their picture (LA2, LA5). Four participants felt that they had no problem recognizing their friends from their profile pictures (LA3, LA7, LA9, LA10). Participants were asked to estimate how many friends on Facebook they had, and participants reported having a mean of 548 (SD = 474) friends. The range of number of friends was from 50 to 1500. However, from talking to participants, recognition of friends seemed to depend less on the number of friends and more on how often they used Facebook. For example, LA2 had difficulties recognizing that the pictures displayed in the proxy were of his friends, and reported that he had not used Facebook in a while. LA5, the participant with 1500 friends, said she had no problem recognizing her friends from profile pictures because she used Facebook a lot and always saw her friends’ latest profile pictures.

The “chattiness” metric received much attention from participants. Two participants specifically reported that it was an interesting and useful way to segment the audience (LA1, LA9). Many participants offered suggestions for other ways to segment the audience, including:

- Segmenting by the topics people chat about (LA3)
- Segmenting by demographics such as age, school, and geographic location (LA8)
- Showing how often people update their Facebook profile (LA9)
- Showing how much “IM slang” people use in chat (LA9)
- Showing who is actively watching vs. who isn’t paying attention (LA10)

Another confusing point about the chattiness metric is that some participants were unclear of whether it was computed for activity “right now” or if it was computed for historical activity. For example, LA7 felt that it could have been based on how much her friends used Facebook; in this case, posting a lot of comments on peoples’ walls constituted “very chatty” behavior. Finally, four participants found it difficult to

interpret their own chattiness rating because they expected it to be represented positionally as it was for everyone else (the interface used a design similar to Figure 11-7c).

Participants were asked about what it meant to have two rings of friends, one ring on the inside the semicircle, and one ring on the outside (shown in Figure 11-8). Many of the participants asked this question had not fully grasped the idea that there were other chat groups in the system, so this question gave particularly good insight into what the rings *could* represent as opposed to what they *did* represent. Six participants generally understood that friends in the inner ring were somehow “closer” (LA3), and that this closeness was defined by some aspect of Facebook activity, such as posting comments or sharing photos (LA6-LA10). The other participants either didn’t know what the rings represented (LA2, LA5), or figured it out after thinking and talking about it (LA1, LA4).

Tag cloud. Participants were split about the tag cloud’s ability to give a good sense of what the audience was talking about. Five participants felt it did give a good sense, and five participants felt it did not. These viewpoints were best summarized by LA6 and LA7.

“Word size emphasizes important topics” (LA7)

“One word doesn’t mean much without other words connected to it” (LA6)

Five participants (including one who liked the tag cloud) said that the tag cloud contained “random words.”

“LOL’ is kind of random” (LA10)

“These are random noises from the audience” (LA8)

Participants felt that the tag cloud could be improved by filtering out irrelevant words (LA5, LA6, LA10), showing original chat messages containing the tag in a tooltip (LA6), using the tags as a way to find other videos (LA6), and letting the user browse the history of the tag cloud to see how it evolved over time (LA4). Almost all participants recognized that the tags came from chat, although LA3 felt that they could have come from the video, such as through its transcript.

Scrolling list. Participants strongly felt that the scrolling list of chat messages gave them a good sense of what the audience was talking about; only two participants disagreed with this sentiment (LA4, LA5). Reflecting the general feeling toward the scrolling list, LA5 said, “This is much better organized even though I don’t recognize the people.” However, she did not feel it was a good summary because she “would get tired because I couldn’t keep six conversations in my head.”

Three participants explicitly mentioned liking the fact that they could see the entire chat message (LA6, LA7, LA10), and two felt that the list could be improved by showing the thread structure of the messages, or by letting the user drill down to see the messages surrounding the ones shown (LA6, LA9). Almost all participants felt that irrelevant messages should not be displayed in the list, and four participants explicitly requested a feature that allowed them to filter messages based on typing in keywords (LA2, LA6, LA7, LA8).

After seeing both the tag cloud and the scrolling list, participants were asked which one they preferred. Seven participants reported preferring the scrolling list, two reported preferring the tag cloud, and one was unsure.

11.5.3. DISCUSSION

Participants liked the ideas of seeing a representation of a large audience of viewers and a summary of their chat, although there were many different interpretations of what exactly constituted a large audience. Even with the anchoring effects of telling participants “we are interested in online video broadcasts with large audiences” and “you are watching this video with about 4.6 million other viewers,” participants had different expectations for the number of viewers it took for an audience to grow from small to large. These varied expectations may make it difficult for collaborative video interfaces to provide viewers with an emotional connection to the audience, such as a sense of awe. If one’s expectation is that a “large” audience is composed of millions of viewers, then he may think that an online event that attracts an audience of 20,000 is small and insignificant, even though the event may be considered successful by its organizers. Thus, it is important for collaborative online video sites to set their members’ expectations. This calibration can be done by showing

historical data for past events, comparing to other similar events, or even just stating how many people were expected to show up²⁵.

Design recommendation 1: Give viewers an anchoring point with which to make a comparison of the current audience size during a live event.

Participants generally felt that the audience proxy gave them a good sense of who was in the audience. However, the proxy was not without its problems. Some participants expressed difficulty in recognizing the faces of all of their friends, either because the pictures were small to discern (as implemented, the pictures were no larger than 24x24 pixels), their friends often changed their profile pictures, their friends' profile pictures were not pictures of them, or participants simply hadn't used Facebook recently. These difficulties can be addressed by adding interaction features to the proxy. Tooltips or separate displays in the interface can show detailed information about each person shown in the proxy. They can also be used to initiate interactions with those people, such as joining their chat group, poking them, making a comment on their wall, or initiating a Facebook private chat. In fact, some of these ideas were explicitly requested by participants.

In addition to difficulties with Facebook profile pictures, participants expressed confusion with the chattiness metric, believing that it was intended to represent historical activity rather than immediate activity. For example, LA7 felt that "chattiness" meant the degree to which her friends made comments on each others' walls. LA3 felt that the friends on the inner ring were closer to him for the same reason – higher mutual activity across time. To address these concerns, the proxy can be configured to display either type of activity, historical or immediate.

One well-received feature of the audience proxy was that it sliced the audience into different facets or groups. Participants offered many suggestions for other ways to segment the audience, including based on their topics of chat, their demographics, or their activity on Facebook. (e.g., posting photos). In fact, this segmentation can be performed based on live feedback from the audience, such as in the case of displaying results from a live poll or quiz. Indeed, the design of the proxy is naturally

²⁵ These numbers can also be approximated in such a way that they enable viewers to make downward comparisons. For example, a viewer who thinks that only 100 people were supposed to show up, and sees that 1000 showed up, might be more awed than a viewer who thinks that 800 were supposed to show up.

suited to displaying the proportional popularity of a small number of categories (Figure 11-6b).

To further support viewers' understanding of the other viewers in the audience, features that allow audience members to interactively and dynamically slice the audience can be added to the social proxy. In this way, viewers can achieve a better understanding of who is in the audience and what their characteristics are. One caveat to making the proxy more interactive is that these interactions require a viewer's attention and input, and thus may be additionally distracting while watching a video. Although more research is needed to determine whether viewers can manage their attention between video, chat, and an interactive social proxy, the incorporation of pausing, intermission periods, or natural down-times during a video (e.g., between plays in a football match) may provide enough opportunity for viewers to interact with the proxy without missing video content.

Design recommendation 2: Audience representations can be interactive and allow users to dynamically explore the composition and characteristics of audience members.

The summaries of chat were received positively and participants understood the need to summarize chat when watching a video in a large audience. Counter to expectations, the tag cloud was not received with high enthusiasm. Tag clouds are popular in online venues such as blogs, photo sharing sites, and bookmarking sites. Feedback from our participants suggests that they may not be as useful for summarizing chat messages, because some of the tags selected were irrelevant. Although Chapter 14 will demonstrate that metrics like IDF can be used to select more relevant words for a tag cloud, participants felt that having the full context of chat messages provided a better summary. They also requested the ability to filter messages on their own, either by using a search feature or a keyword whitelist.

Despite the mixed preference for tag clouds and the strong preference for the scrolling list of messages, we recognize that these options are not mutually exclusive. Both can be supported in a collaborative online video interface, and users can choose to display the visualization most appealing to them; indeed, this is the strategy employed in the interface presented in Chapter 12. However, as with the audience proxy, the summaries can be improved by allowing users to interact with them. Tag clouds can show individual chat messages when the user clicks or hovers over a tag. The scrolling list can highlight popular or important keywords in

individual messages. It can also provide more context for each message by showing the messages that came before or after it, either flat or hierarchically²⁶ in its group. Both summaries can be used as a basis for finding chat groups to join, and both summaries can be searched or filtered with keywords. The design space for visual, interactive chat summaries is vast, and this work only scratches the surface of what can be done.

Design recommendation 3: Provide multiple chat summaries to viewers and allow them to monitor the one most interesting to them. Provide access to the full content and context of individual chat messages, and allow users to specify topics of interest to filter out irrelevant messages. Enable chat messages to act as a gateway to items such as other videos to watch or other users with whom to chat.

11.6. GENERAL DISCUSSION

This chapter discusses the challenges present in designing an enjoyable, collaborative online video watching experience for an audience of millions of simultaneous viewers. In this case, the traditional model of chat explored in Part II cannot be directly applied; text chat rooms have a natural limit to the number of people they can support (Jones et al., 2008). Therefore, the audience is segmented into a number of smaller chat groups. The experience of participating in these smaller chat groups should be similar to that studied in Part II, albeit with the understanding that behaviors observed in laboratory studies do not necessarily reflect behaviors seen in the real world. Nonetheless, for those viewers interested in chatting with other members in the audience, the chat group metaphor provides a space to hold back-and-forth conversation. Other metaphors, such as aggregating-and-filtering chat messages, make back-and-forth conversation difficult, though not impossible²⁷.

This chapter discussed a prototype user interface with two important features for watching live video in a large audience: a representation of that audience using a novel social proxy, and a visual summary of the chat topics or messages occurring in

²⁶ Smith, Cadiz, and Burkhalter (2000) observed that people had difficulties in following conversations when presented hierarchically, although those difficulties stemmed from the fact that new content was not always added to the same place.

²⁷ Back-and-forth chat is possible, although difficult, in an aggregate-and-filter model. Common practice in Twitter, for example, is to place an “@<recipient>” tag before a message to specify its intended recipient.

other groups. Participants in the Large Audience study liked the idea of representing the audience. They generally came to understand the audience proxy, either on their own or by talking it through, and felt that it gave them a good sense of who else was in the audience. The visual summaries of chat – tag clouds and a scrolling list of messages – were also liked, although there was a stronger preference for the scrolling list of messages because it displayed the full contents of the chat messages, and not just “random words.” Participants offered many suggestions for improving the audience proxy, the tag cloud, and the scrolling list. These suggestions included requests for more interactivity with the interface, such as linking information between the different chat summaries and using the audience proxy as a way to initiate interactions with other viewers. Finally, participants’ confusion as to whether the audience proxy was showing them current or historical behavior motivates the transformation of the audience proxy into a more generalized “social dashboard” that summarizes a user, their friends, and their mutual activities on Facebook. This notion is discussed further in Section 17.5.

Participants in this study were asked about the types of chat groups in which they would be most interested. Overwhelmingly, they reported preferring groups with their friends, although some participants were interested in chatting with strangers, either alone or in the company of their friends.

Participants also strongly felt that watching video with other viewers who were synchronized in their video playback was important. This finding supports the notion that chat may ruin the indeterminacy of a video when other viewers see an important event a few seconds ahead and spoil the moment by chatting about it (Vosgerau, Wertenbroch, & Carmon, 2006). Note that this type of synchronization is not the same type of synchronization discussed in Chapter 9; in that case, the unsynchronized viewers were not meant to watch the same videos at the same time, and the videos they watched did not necessarily have the same indeterminacy requirements as, for example, sporting events.

The strong preferences for being synchronized with the other viewers in the chat group, chatting in small groups (2-20 people), and watching with friends, motivates a relaxation of the strong synchronization goals typical of peer-to-peer video streaming protocols. Overlay tree protocols like End System Multicast (Chu et al., 2001; Chu et al., 2004) and SplitStream (Castro et al., 2003), and mesh-based protocols like Chainsaw (Pai et al., 2005), strive to keep all viewers’ video playback

synchronized with each other. They do this by optimizing their trees or meshes according to the available bandwidth and the latencies of each node. Nodes with more bandwidth have a greater capacity to deliver video data to more viewers, and nodes with lower latency are clustered together to keep viewers synchronized. These protocols often strive to make a strong synchronization guarantee by minimizing total delay across all viewers (Brosh, Levin, & Shavitt, 2007). From the perspective of an individual viewer, this minimum-delay guarantee translates to the result that all viewers are watching the video within N seconds of each other. Lower values of N correspond to stronger degrees of global synchronization. Implementing a strong synchronization guarantee is difficult and trades-off with the guaranteed delivery of video data. For example, a system concerned with strong synchronization will not spend as much effort retransmitting lost data packets, resulting in choppy video, because it prioritizes the newer packets that keep viewers synchronized.

In light of the preferences for chatting in small groups with friends, strong synchronization among *all* viewers may not be necessary. If we think of each chat group as a cluster of nodes in the system, we only need to ensure strong (e.g., sub-second) within-cluster synchronization. Because viewers prefer to watch with friends, and because they prefer watching in small groups, the size of these clusters would be small (e.g., < 20 people), and there would be an increased likelihood that the nodes within each cluster are geographically close to one another. Thus, assuming viewers' behavior corresponds to their reported preferences, this new protocol can focus on providing strong (sub-second) synchronization guarantees among viewers within a cluster, and weaker synchronization guarantees between clusters (e.g., 3-10 seconds). This weakening of the synchronization requirement enables nodes to reduce their packet loss rates by transmitting more data over higher-latency links, resulting in less-choppy video for all viewers.

This idea adds the *social network* as a new resource that can be leveraged when designing video streaming protocols, in addition to point-to-point latency and bandwidth metrics. The social network provides an opportunity for relaxing video synchronization requirements – one of the most difficult requirements to meet – of peer-to-peer video streaming systems.

11.7. SUMMARY AND CONCLUSIONS

- Many live online video events are attracting large-scale audiences of millions of simultaneous viewers.
- Two methods for enabling audience members to chat and share messages with each other are discussed. Aggregation-and-filter, used by the Facebook Live Chat Box and Twistori, shows chat/status messages over time but does not promote back-and-forth conversation. Small chat groups promote back-and-forth conversation but do not provide information about the activity of the entire audience.
- This chapter studies a hybrid approach that uses a combination of tag clouds, a scrolling list of messages, and a novel social proxy representation of the audience to provide awareness of the entire audience and to summarize their chat messages with each other. A think-aloud user study collected feedback on these features. Participants liked the ideas of representing the audience and summarizing their chat.
- The audience proxy, modeled after a radar screen, encodes continuous attributes of audience members using angular position and distance from center. Participants generally came to understand the audience proxy and felt that the ability to slice the audience provided them with a good sense of who else was in the audience.
- In addition to chattiness, participants expressed interest in other ways of slicing the audience, such as based on their age, their geographic location, or their use of “IM slang”.
- Tag clouds were not rated as useful as a scrolling list of messages because of the presence of irrelevant words. Word filters can be used to improve the relevance of messages and tags displayed in the chat summaries.
- Audience and chat summaries should be interactive and allow users to initiate interactions with audience members, reveal additional context of chat messages, and find chat groups to join. Future work is needed to determine the extent to which interactive visualizations are distracting in the collaborative watching context.
- Participants’ expectations of what constitutes a “large” audience were greatly varied. Event organizers can display an expected or historical

audience size to anchor participants' expectations and help them understand the magnitude of the audience's size.

- Participants reported preferring watching videos in small groups of their friends. This observation motivates the use of social networks as a resource for P2P video streaming protocols to reduce their overall loss rates. Future work is needed to quantify the benefits of a protocol that utilizes social network data to optimize data delivery.

12.

THE “SOCIAL VIDEO” APPLICATION

The capstone of this dissertation was the design, implementation, and deployment of a real-world collaborative online video watching application called “Social Video”. The motivations behind creating this application were twofold: first, it allowed for the observation of real-world behaviors of viewers watching videos online; second, it provided an opportunity to design a collaborative online video application informed by the findings of the studies presented in this dissertation.

12.1. SCENARIOS

Social Video was designed and implemented by myself and several other students. In designing it, we constructed the following scenarios to guide our thinking. These scenarios were based on our own experiences watching videos online and was informed by the results from the studies discussed in Part II.

Keeping in touch with friends.

Courtney just started college and misses her high school friends. She wants to keep in touch with her friends so she decides to make a list of videos that she and her friends would enjoy watching together in the Social Video application.

Courtney wants to prepare the chat group with her videos before she invites her friends. She creates a new chat group, searches YouTube for the videos on her list, and adds them to the chat group’s shared playlist.

After she has added all of her videos to the chat group, she invites her friends Diana and Teresa to watch videos with her. Once Diana and Teresa log on, they see the invitation to Courtney’s chat group and join it. The three friends catch up with each other as they watch the videos together.

Maintaining a long distance relationship.

John and Jane are in a long distance relationship. They are looking for an online activity they can share to keep the relationship going while they are apart. They decide to watch movies together online every Saturday night.

It is Saturday night, close to movie-watching time. John and Jane both visit the Social Video site. As this is not the first time they have watched a movie together, they see the chat room they created for watching movies together. They click on the room to join it.

John and Jane never decide what movie to watch beforehand. When they join the room, they use the Find Video feature to search for a movie to watch. Jane sees several interesting movies and adds them to the chat group’s shared playlist. She also scans the list of movies they have previously watched to make sure the new movies she added haven’t been watched yet. Jane and John chat about which movie they want to watch and decide on one of them. Jane moves that movie to the top of the list and clicks “Start Next Video” to begin playing the movie.

Jane and John watch the movie and chat with each other. When playback is finished, they see a waiting screen that shows that they both have finished watching the movie. They continue chatting for several minutes after the movie has finished. When they say goodbye, Jane closes the window, confident that the next time she joins the system, she will be able to join her chat group again.

12.2. DESIGN GOALS

The Social Video application was designed to satisfy the following goals.

Goal 1: Provide a fun and sociable experience for all online video

Chatting while watching videos is fun. The Social Video interface was designed to act as a wrapper for videos from any online source. This wrapper included a text chat feature, which enabled viewers to interact with each other while watching together.

Goal 2: Support watching with friends

Current online video sites do not make it easy for friends to watch videos together because they do not know who one’s friends are. As seen in Part II, watching with friends is enjoyable and has social benefits. A study by Joinson (2008) showed that friends largely used Facebook to keep in touch with each other. Thus, Social Video took advantage of the Facebook platform to tap into viewers’ real-world social networks and make it easy for friends to watch videos together.

Goal 3: Encourage encounters with strangers

In a laboratory context, we observed that watching and chatting with strangers led to increased feelings of liking and closeness with those strangers, and strangers

talked about personal topics with each other (Chapter 8). However, the survey responses discussed in Chapter 11 indicate that people were more interested in watching videos with their friends than strangers. Further, other research shows that there is a general disinterest in meeting new people on Facebook (Lampe, Ellison, & Steinfield, 2006). Therefore, online communities that wish to encourage bonding among members through social interaction while watching videos may face significant resistance from users who simply wish to interact with their friends.

To encourage encounters with strangers in our system, chat took place among groups of people, rather than in dyads (such as in Zync). Viewers could also join multiple chat groups simultaneously so they could explore alternative chat groups without losing their membership in the groups they enjoy (i.e., the groups with their friends). Visualizations of chat and video-watching activity over time were also used to make the activities of strangers visible. These methods were used to increase the transparency of activity in the system and encourage “stranger encounters” among viewers. As will be discussed in Section 12.5, we were unable to collect enough data to adequately determine whether these mechanisms were effective.

Goal 4: Incorporate large audience features

In Chapter 11, I presented the design of a social proxy for representing the composition and activities of a large audience, as well as visualizations that summarize their chat messages. People generally found these designs interesting and useful. Thus, we included these features in our design.

Goal 5: Maintain ecological validity

A study is ecologically valid when its methods, materials, and setting approximate the real-life situation under investigation. For Social Video to be ecologically valid, it needed to approximate the feature set and mode of operation of current collaborative online video sites. This goal was met by allowing viewers to watch videos from current online video sites, by hooking into existing real-world social networks on Facebook, and by using a text chat feature common to many collaborative video sites.

Goal 6: Implement as few control/authority mechanisms as possible

We took a relaxed stance on implementing control and authority mechanisms in Social Video to understand if coordination could be performed among viewers. Thus,

there were no group-level controls for managing chat groups (i.e., muting people in a group). As a consequence, the implementation task was easier (fewer features to implement), but more importantly, this policy provided an opportunity to understand whether control/authority features were really necessary in a setting in which individuals were identified by their real-world identities. Both Siegel et al. (1986) and Kiesler and Sproull (1992) found that computer-mediated communication reduces people’s inhibitions toward behaving in an anti-social manner, resulting in “flaming” behaviors. Although we speculated that “flaming” behaviors would not often occur among non-anonymous users, application usage was not great enough to adequately draw a conclusion.

12.3. DESIGN DECISIONS

In Chapter 3, I presented a framework for the design of collaborative online video experiences. These decisions revolve around the video content, the video distribution technology, the viewing device, the model of video playback, and the nature of social interactions supported.

Video content. Social Video was designed to act as a “social wrapper” around any video content online (Goal 1). However, when implementing the application, we found that embedding videos from online video sites was subject to the embedding policies of those sites and, in many cases, required a per-site implementation. For example, some sites provided APIs to access videos and their metadata (e.g., YouTube), whereas other sites (e.g., Hulu) did not. Thus, the implementation focused solely on watching videos from YouTube to maintain ecological validity (Goal 5).

Viewing device. Social Video was designed for use on computers by viewers who were not physically co-located. The design was not tailored for the stricter user interface requirements of the television.

Playback model. The studies discussed in Part II suggest that for groups of friends, watching the same videos at the same time is as enjoyable as watching the videos in a different order. For strangers, it is unclear if the lack of common ground when watching different videos would discourage chat. Other collaborative online video systems have studied the synchronous case in which viewers watch the same videos at the same time (Nathan et al., 2008; Shamma et al., 2008; Liu et al., 2007). Thus, in keeping with the goals of ecological validity (Goal 5) and encouraging encounters

with strangers (Goal 3), Social video employed a hybrid model of video playback. Viewers in the same chat group watched the same videos at the same time. These videos were drawn from a shared, per-group playlist that anyone in the group could modify by adding, removing, or reordering videos.

An alternative to the per-group playlist system is one in which chat groups are created dynamically based on who else is watching the video at the same time. This design is video-centric and is akin to combining real-time chat with the also-watching feature on YouTube. This approach is less ideal than the group-centric approach for promoting sociability for friends and strangers. Many videos on YouTube are short, typically lasting between three to five minutes. This amount of time is short for chatting with other people, and thus discourages back-and-forth conversation. In addition, if users are migrated between chat sessions when their video changes, their conversations may be cut short when a video finishes playing. Finally, video-centric groups would require a feature where friends could stick together and follow each other through a series of videos to support watching videos with friends. Therefore, chatting in Social Video was group-centric.

Distribution technology. Since Social Video played videos embedded from other online video sites, it did not need its own video distribution system. However, the capability of incorporating other distribution mechanisms into the video player did exist (as long as those mechanisms were implemented in Adobe Flex).

Interaction. Many collaborative online video sites allow viewers to watch videos and post messages or comments about videos in an asynchronous fashion, such as YouTube. Others have viewers watch the same videos together and provide a real-time, synchronous chat feature for viewers, such as UStream.TV. Thus, either choice would have been ecologically valid. The results from Part II show that real-time chat with synchronized video provides an enjoyable experience for both friends and strangers. Thus, Social Video had remote viewers – both friends and strangers – watch videos at the same time, and in the same “place” of a chat group. Viewers interacted with each other using a text chat feature.

12.4. IMPLEMENTATION

The Social Video application was implemented using a combination of Adobe Flex, Facebook Markup Language (FBML), and Facebook Javascript (FBJS). The back-end

server was implemented with PHP and MySQL. Figure 12-1 shows a screenshot from the initial implementation of Social Video²⁸. In Section 12.7, I present a redesigned user interface based on usability testing and feedback from the deployment trial.

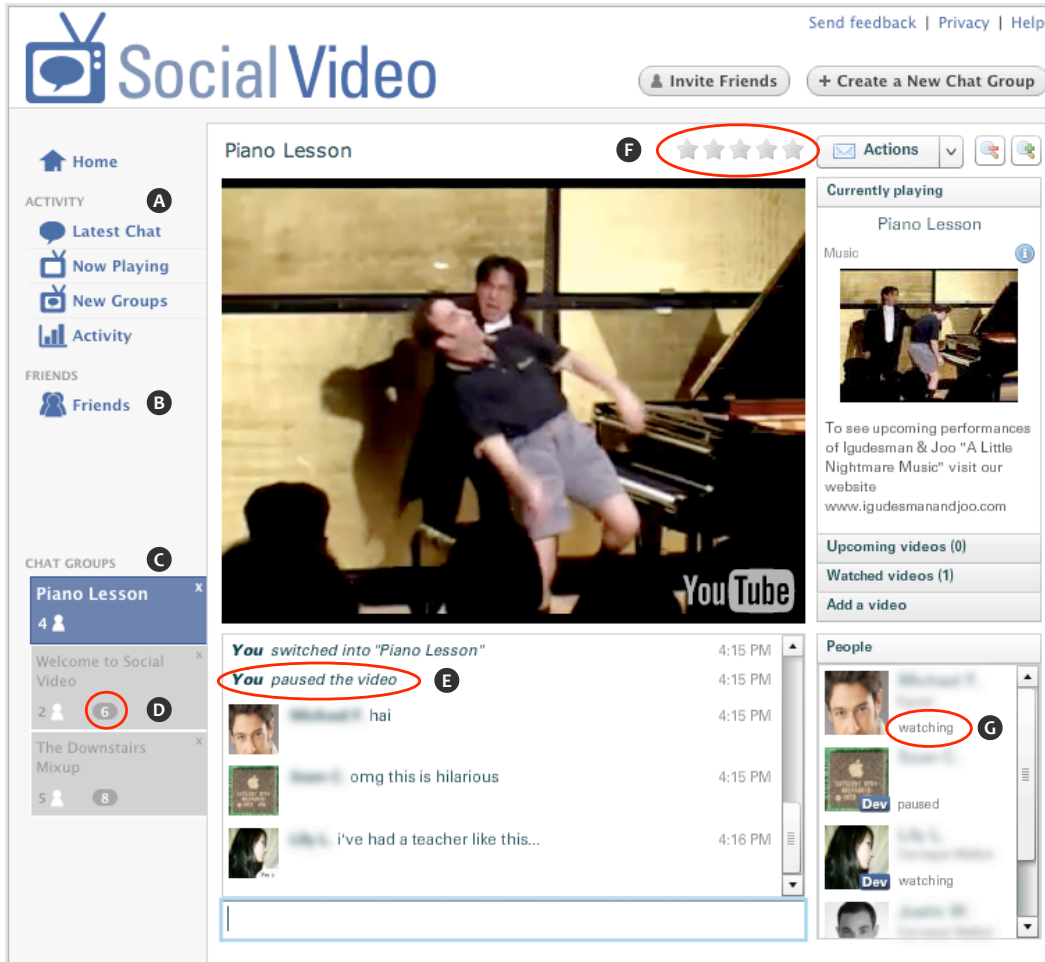


Figure 12-1. The Social Video application. This interface was situated on the Facebook canvas page. Advertisements and Facebook toolbars have been removed. Names have been blurred for anonymity. (a) Visualizations of system activity. (b) List of online friends. (c) Multiple persistent chat groups can be joined. (d) Unread message counts allow users to monitor activity across their chat groups. (e) Status messages in chat display users’ playback activity. (f) Group ratings help people find interesting groups. (g) Video playback status is visible for all users.

In the rest of this section, I discuss the specific features we implemented to satisfy the design goals and give a sense of how this application operated.

²⁸ For historical reference, Social Video was available at <http://apps.facebook.com/social-video>. It is highly unlikely that this application is still available for public use.

12.4.1. PER-GROUP SHARED PLAYLISTS

Each chat group contained a playlist of videos, shown in Figure 12-2. The playlist was a queue of videos that would be watched by the entire group. It was a shared playlist for all members of the group, rather than a set of individual playlists for each member of the group. Anyone could add videos to, remove videos from, or reorder videos in the playlist. Videos were played in a first-in, first-out order. The initial implementation of the playlist only supported adding videos; support for removing and reordering videos was implemented later.

In a set of design guidelines for social television systems, Ducheneaut et al. (2008) recommend that social television systems provide a preview of the oncoming show structure. The per-group playlist provided group members with this preview.

12.4.2. INITIALLY-SYNCHRONIZED VIDEO

Video playback was synchronized among viewers so they watched the same videos at the same time. However, this synchronization only occurred once per video, when the video player received the signal to begin playing the video. This signal was sent when a user clicked the “Start Next Video” button (Figure 12-2). Any viewer in a chat group could press this button. This button’s implementation contained logic to ensure that it was only triggered once when multiple viewers clicked it in close temporal proximity.

This method of video synchronization is similar to how Zync synchronizes video among dyadic viewers (Shamma et al., 2008). In both systems, a video start signal is used to notify all viewers to begin video playback. However, unlike Zync, our system did not synchronize any subsequent pause, play, or seek signals. We call this relaxed form of synchronization “initially synchronized” because viewers are synchronized only when a new video should begin playing. Viewers may then become desynchronized for several reasons.

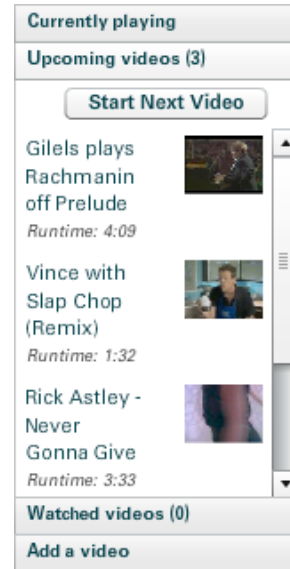


Figure 12-2. The shared video playlist. The “Start Next Video” button caused viewers to immediately start playing the next video in the playlist.

- Buffering time and lag cause variance in the time at which a viewer begins video playback,
- Network jitter causes video to stutter during playback,
- High network latencies, congestion, and dropped packets require a viewer to perform additional buffering, and
- Viewers may pause, resume, or seek within their own video, which immediately desynchronizes them from other viewers.

The first three reasons are harsh realities of the Internet. Although video synchronization and distribution protocols exist to cope with latency, congestion, and buffering times (e.g., Schulzrinne et al., 1996; Chu et al., 2004), our application did not use them. The last reason was motivated by Goal 6 (implement as few control structures as possible), whereby the pause feature only paused the video for the individual who pressed pause, instead of pausing for the entire group.

The decision to use initially-synchronized video was made from two data points. First, the Text vs. Audio study (Chapter 9) showed that friends enjoyed chatting while watching unsynchronized content. Second, during the design process, one of our potential users explained a pitfall of having strong video synchronization. He said that he would not want to have to wait to watch a video because his friend was on a poorer Internet connection and required a longer amount of time to buffer the video. Instead, he would want to begin watching immediately and be notified when his friend had begun playback. In chat groups with dozens of people, the distribution of Internet connections is likely to be broad and the variance in latencies to the source of the video is likely to be high. If viewers in a chat group were strictly synchronized, there would be situations in which people would be stuck waiting for one person to finish buffering a video before anyone could begin playing it. Waiting for one friend to buffer a video may be acceptable; waiting for ten of them to do so may be frustrating.

To help viewers understand the playback state of the other members of their group, we relied on the principle of social translucency (Erickson & Kellogg, 2000) by making each viewer’s playback status – whether they are buffering, playing, paused, or finished watching a video – visible (Figure 12-1g). This information allowed viewers to make an informed decision about when to begin watching the next video together, which also had the consequence of re-synchronizing the group.

One question that arose during the implementation of initially-synchronized video was about what happened when a viewer’s video finished playing? Because viewers started and finished watching a video at slightly different times, viewers needed something to look at while they waited for everyone else to finish. Again, relying on guidance from the laboratory studies, we decided to use the lack of strong synchronization to our advantage. The Cartoon study showed that brief intermission periods between videos helped viewers feel less distracted from the chat feature. Therefore, we allowed viewers to create intermission periods between videos by showing them a waiting screen when their playback finished (Figure 12-3). This screen provided viewers an opportunity to chat without being distracted, and it provided another spot in the interface to make viewers’ playback status visible.

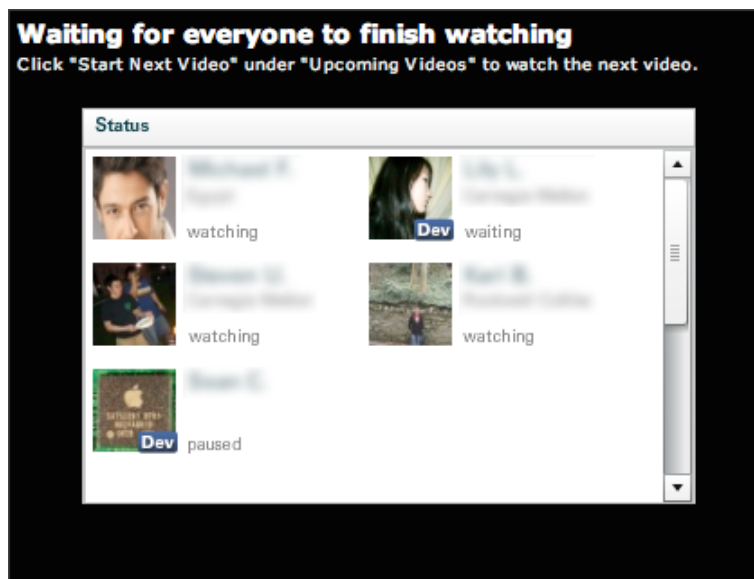


Figure 12-3. The waiting screen. This screen was shown to viewers once the current video finished playing for them. The playback status of each person in the group was shown. Names have been blurred to preserve anonymity.

To further increase the translucency of activity in our system, messages were displayed in the chat log when people joined or left the group, added videos to the playlist, started the next video, or changed their playback status. (Figure 12-1g). These messages further aided their coordination.

Ducheneaut et al. (2008) recommend that social television systems automatically sense when viewers are conversing and adjust video playback accordingly. For example, when a group’s discussion during a commercial break is intense, the system can insert additional commercials to give the group time to finish their

discussion before they resume watching the television show. In our system, the waiting screen simulates this process, albeit manually: group members themselves decide when to watch the next video, not the system.

12.4.3. PUBLIC AND PRIVATE CHAT GROUPS

Chat groups in Social Video could either be either public or private. Public groups were advertised in the system and on Facebook, and private groups were not. To promote encounters with strangers, chat groups were public by default (Goal 3). However, because friends may have wanted to create places just for themselves, we implemented private chat groups as well, with membership by invitation only (Goal 2). Chat groups could be made by any user, and users could make as many groups as they liked (Goal 6).

One benefit of having both public and private groups is that they provided behavioral insight into whether people were interested in encountering new people, or if they simply wanted to watch videos with the friends they already had. The survey results discussed in Chapter 11 suggest that people generally only want to watch videos with their friends. If the majority of created groups were private, then we find behavioral evidence supporting this result. However, if the majority of created groups were public, we might infer that people were open to interactions with strangers. The distribution of private and public chat groups is discussed in Section 12.5.

12.4.4. PERSISTENT CHAT GROUPS

In designing the chat group feature, we had to decide whether to make groups persistent, by keeping them in the system even after all members had left, or ephemeral, by removing them once all members had left. There are interesting tradeoffs between group persistence, supporting the needs of friends, and encouraging encounters with strangers. For groups of friends, persistent groups can become virtual “places” they go to hang out each time they use the application (Dourish, 2006). Persistent groups also enable friends to schedule a viewing session ahead of time by creating a group and populating it with videos. For example, Courtney (described in the scenario) wants to watch videos with her friends, and she wants to choose videos beforehand. With persistent groups, she can create a

chat group, add videos to its playlist, send invitations to her friends, and close her web browser until it is time to actually watch those videos. With ephemeral groups, she would have to keep her browser window open in order to keep the group alive.

Although persistent groups have advantages for friends, they may reduce peoples’ opportunities to interact with strangers. If people habitually join the same groups each time they visit the application, they may be discouraged from actively seeking out new groups to join. Ephemeral groups encourage users to explore and meet new people, because each time they use the application, they are presented with new options for groups to join.

We chose to make groups persistent in our application, because we felt the benefits to groups of friends outweighed the risk of discouraging encounters with strangers. On the surface, our application could have supported both types of groups by asking users to choose the persistence model when creating a group. However, a system with both persistent and ephemeral groups still has persistent groups, and thus does not gain the advantages afforded by ephemerality.

When a user joined a chat group, that group was displayed in their chat group list (Figure 12-1c). Chat group memberships were restored each time the user logged into the application, so they did not have to find the same groups over and over again. Ducheneaut et al. (2008) recommend that social television systems make it easy for viewers to move in and out of the audience smoothly, and our interface made it easy for users to switch among chat groups by clicking the appropriate tab.

12.4.5. GROUP INVITATIONS AND FINDING FRIENDS

To make it easy for friends to watch videos together (Goal 2), they could send invitations to chat groups to each other. Invitations could be sent inside the application, which placed a link to the chat group on the home page of the invitee. They could also be sent out-of-band using a URL created specifically for each group. Further, friends could find each other when they were online by loading the Friends page (link shown in Figure 12-1b).

12.4.6. FINDING CHAT GROUPS

Our approach to supporting a large audience of viewers was to fragment that audience into many, smaller chat groups, as discussed in Chapter 11. This fragmentation was independent of the videos viewers watched; thus, if thousands of people watched the same video at the same time (e.g., a live video feed), they could have done so in one of hundreds of different chat groups. However, in this situation, viewers have a choice problem: with too many chat group options, which group should they pick, and how could they even compare different groups?

One strategy for helping users find interesting groups when many are available is to use a recommender system. Chapter 11 provides some insight into how to design such a recommender system. To supplement that insight, we collected behavioral data on the types of groups viewers enjoyed by having viewers rate the chat groups in which they participated (Figure 12-1f). These ratings help us understand why people enjoyed a chat group – because of the videos they watched, or the social experience they had. In Section 12.6, I discuss an evaluation of the ‘features’ of a chat group that predict one’s enjoyment of that group.

To make it easy for new users to find groups to join, we employed a simpler strategy for recommending groups. On the home page of each user, we listed a random set of popular groups (those with active members). This way, even when the system had only a few users online, other users could easily be found.

12.4.7. MAINTAINING COMMUNITY AWARENESS

The initial version of Social Video incorporated the tag cloud and scrolling list visualizations, discussed in Chapter 11, for summarizing users’ chat and video-watching activities in the system. As the development of the audience proxy concept was concurrent with the implementation of Social Video, this feature was not included in the initial deployment.

The “latest chat” visualization (link shown in Figure 12-1a) was an implementation of the scrolling list visualization. It displayed chat messages from public chat groups, along with a link to the group in which the chat message was spoken. In the implementation, chat messages were selected randomly, with a preference given to more recent messages. Figure 12-4 shows a screenshot from this visualization.

Latest chat

Here is what people are talking about on Social Video. Chat messages are randomly picked from all chat groups (except the private ones).



Figure 12-4. Latest chat visualization. Chat messages were randomly selected from public chat groups, with a preference for newer messages. Names have been blurred to protect anonymity.

Similar visualizations were used to display the videos that chat groups were currently watching, as well as chat groups that had recently been created. These visualizations were accessed through the “Now Playing” and “New Groups” links shown in Figure 12-1a.

We provided additional visual summaries of system activity in the form of a tag cloud of popular chat terms and a tree-map of popularly-watched videos. Due to time constraints, the initial implementation of these features were display-only summaries. However, both summaries were intended to be a gateway for finding groups through their topics of chat and the videos they were watching (following design recommendation 3 from Chapter 11). For example, we intended the terms in the tag cloud to be used for finding groups that talked about those topics. We also intended the videos in the tree-map to be used for finding groups that were watching (or had queued) those videos. Linking videos to groups that are currently

watching them is one way a group-centric system can help users find alternative social experiences for a given video.

12.5. GENERAL USAGE

The Social Video application was publicly deployed on Facebook on April 29, 2009. As this application collected data as part of a research project, three requirements were made for using this application: users had to have a Facebook account, users had to be 18 or older, and users had to sign an online consent form to participate in our study; the last two requirements were to maintain compliance with our IRB.

To receive early feedback for iterative development and bootstrap our user base, we recruited our initial users from several sources, including our own social groups, fliers posted around campus, and messages posted to a popular campus electronic board (*misc.market*). We also recruited users for an evaluation study from Amazon’s Mechanical Turk, discussed in the next section. One consequence of our methods of recruitment are that we attracted enough small groups of friends to draw conclusions about general enjoyment and usage (Goals 1 & 2), but we did not reach enough of a critical mass to evaluate the goal of promoting encounters with strangers (Goal 3), or to study the effect of minimizing authority mechanisms (Goal 6). Evaluation of the features that supported large audiences (Goal 4) was conducted in the laboratory study discussed in Section 11.5.

Figure 12-5 shows the number of users who registered to use the application over time. Use of the application began with beta testing on March 30. The evaluation study ran from May 13 to May 26. Usage decayed over the summer and ceased at about mid-September primarily because of our focus on redesigning the interface and making compatibility improvements, rather than marketing the application to additional audiences.

General usage was lower than expected due to our inability to reach a critical mass of users. One challenge to widespread adoption faced by our application is that the social features required other users to be present, at the same time, in order to be useful. Thus, a user who visited the site and found no one else there may not have stuck around for long themselves; when this process was experienced by many users (excluding those who visited the site with friends in tow), it led to a situation in which the site may have felt abandoned or unused. Using a blend of synchronous

and asynchronous interaction features may have been helpful in avoiding this particular situation.

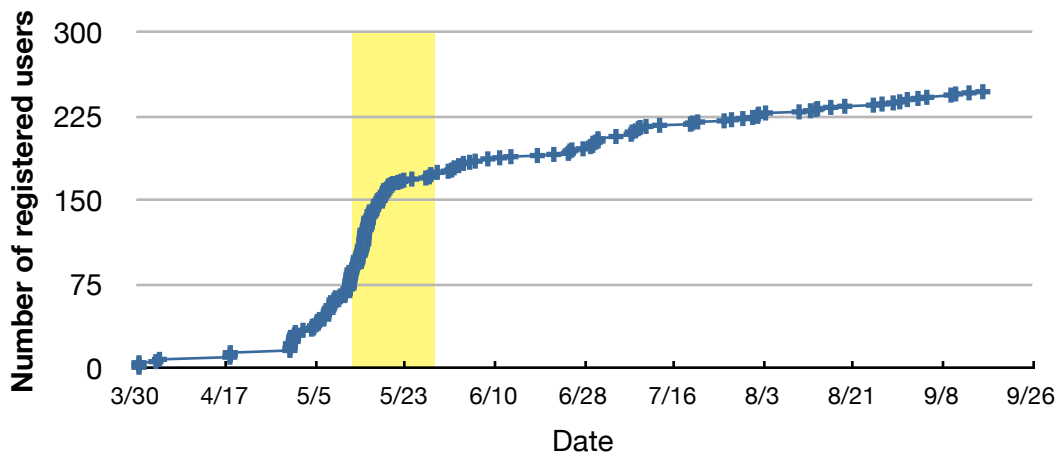


Figure 12-5. Number of registered users over time. Overall, 247 users signed up to use the application. The highlighted period from May 13 to May 26 corresponds to the evaluation study discussed in Section 12.6.

As for the usage the application did receive, users created 37 chat groups during the usage period (this number includes several groups created by the developers specifically to attract new users). Of these, only 5 (13.5%) were private. Only 14 (37.8%) chat groups had more than one person join, suggesting that many groups were created by users just to test out the system. The public groups created by the developers, including the introductory “Welcome to Social Video” group, attracted more users.

Although the total number of groups created by users was quite low, the relative infrequency of private groups suggests either that users did not mind having their groups advertised in the system and joinable by other users, or that users simply accepted the default value for creating public groups.

12.6. EVALUATION STUDY

We performed an evaluation study of the initial version of our application to gain feedback on the design, to promote the application’s usage, and to observe real-world patterns of watching and chatting for different types of videos.

12.6.1. RESEARCH QUESTIONS

The Social Video study was an observational study designed to address the following questions related to watching and chatting in a real-world context.

The first three questions relate to how viewers took control over their watching and chatting experience. In the Cartoon study, intermission periods were placed between videos, and viewers could not shorten, lengthen, or skip these periods. In Social Video, viewers can create break periods for themselves in two ways: they can all agree to pause the video together, or they can spend time chatting on the waiting screen before starting the next video. With control over video playback, will they confine their chat to just the break periods?

RQ 12-1: How do viewers manage the activities of chatting and watching when they are in control of their own video playback?

RQ 12-2: Do viewers take breaks between videos to chat without being distracted?

RQ 12-3: Do viewers use the pause feature to create their own break periods?

The last question is about understanding which aspects of the experience affect viewers’ enjoyment more: the videos they watch or the amount they chat.

RQ 12-4: How does viewers’ enjoyment relate to the videos they watch and the amount they chat?

12.6.2. PARTICIPANTS

Amazon’s Mechanical Turk²⁹ is a site that allows users to post Human Intelligence Tasks (called “HITs”), which are accepted and performed by “workers”. Workers are compensated for their time and quality of work. Prior work by Kittur, Chi, and Suh (2008) has found Mechanical Turk to be effective for conducting user studies that require participants to complete micro-tasks, although care needs to be taken to detect users who attempt to game the system.

We posted a total of 32 HITs on Mechanical Turk that required workers to watch a series of videos with two or more of their friends. Five HITs were removed from

²⁹ Amazon Mechanical Turk. <http://mturk.com>

analysis because the workers did not watch the videos, they did not watch with friends, they experienced technical difficulties that prevented them from completing the task, or no workers accepted the task. Data in this study are from 27 video-watching sessions conducted by 20 distinct groups of friends. As five friend groups participated in multiple sessions, each distinct group of friends is assigned a unique ID that is used as a random factor in the analysis.

12.6.3. METHOD

Workers were instructed to find two or more friends with whom to watch videos. They were given a link to a pre-created, private chat group in Social Video that had the videos set up ahead of time in the playlist. Participants received instructions on how to start the next video in the HIT instructions, and were also told to rate each video after it had finished, and to rate the chat group after all videos had finished.

Videos of several different genres were used to appeal to a wide variety of tastes, and also to compare between two classes of videos: informational and entertainment. This distinction stemmed from an observation made by Geerts, Cesar, and Bulterman (2008) that peoples’ preferences for talking during content depended on its genre. Videos of differing lengths were used as well; in some cases, longer videos were employed to examine collaborative watching with longer content, and in other cases, sets of shorter videos were used to examine whether participants created their own break periods. Each set of videos lasted between 30 and 40 minutes, and participants were paid \$9 for their time. Table 12-1 details the sets of videos used in this study.

Table 12-1. Videos used in the evaluation study. The number of videos in each set and the number of HIT groups that watched those videos are given.

Video set	Description	# Videos	# Sessions
Informational			
Business	Interviews with business leaders on growth and innovation	3	4
Robots	Lecture on ethics in robotics	1	4
Pittsburghese	Lecture on the Pittsburghese dialect	1	3
Entertainment			
Music	Clips of musical performances and compositions	10	3

Video set	Description	# Videos	# Sessions
Starcraft	Video game tournament with commentary	1	3
Red vs. Blue	Episodes from a popular video game machinima series	5	3
Ultimate	Ultimate frisbee match with commentary	3	4
Sports	Sports clips from baseball, basketball, soccer, hockey, and tennis	6	3

12.6.4. MEASURES

Two primary measures were used in this study. Behavioral measures of when participants watched videos, when they paused the videos, and when they chatted were used to address the first three research questions about how viewers manage their attention between chatting and watching when they control video playback.

Several self-reported measures were used gauge enjoyment. Participants were asked to rate the HIT upon completion – “How much did you enjoy doing this HIT?” – on a 5-point open-ended scale anchored by “Not at all” and “Very much”. Participants (and their friends) were also asked to rate their chat group using the 5-point star scale (1 to 5 graphical stars) inside the application. To remove a ceiling effect in the HIT ratings, as well as compute ratings for groups that either did not rate the HIT or did not rate the chat group, we operationalized enjoyment as the average of all of these ratings. Higher numbers indicate greater enjoyment.

Participants were also asked to rate each video immediately after they finished watching it to assess enjoyment of the video content. As with the ratings of the chat groups, the videos were rated on a 5-point star scale.

12.6.5. RESULTS

Patterns of watching and chatting. The first two research questions are about how viewers managed their time between watching videos and chatting. Figure 12-6 shows a timeline of the activity in each viewing session. For each session, colored bands show when a video was being watched; the color indicates which set of videos were watched in that session. Gaps between these bands indicate when the waiting screen was displayed. Black dots mark when participants sent chat messages to each other. For interpretability, and to be able to compare patterns across groups, time is

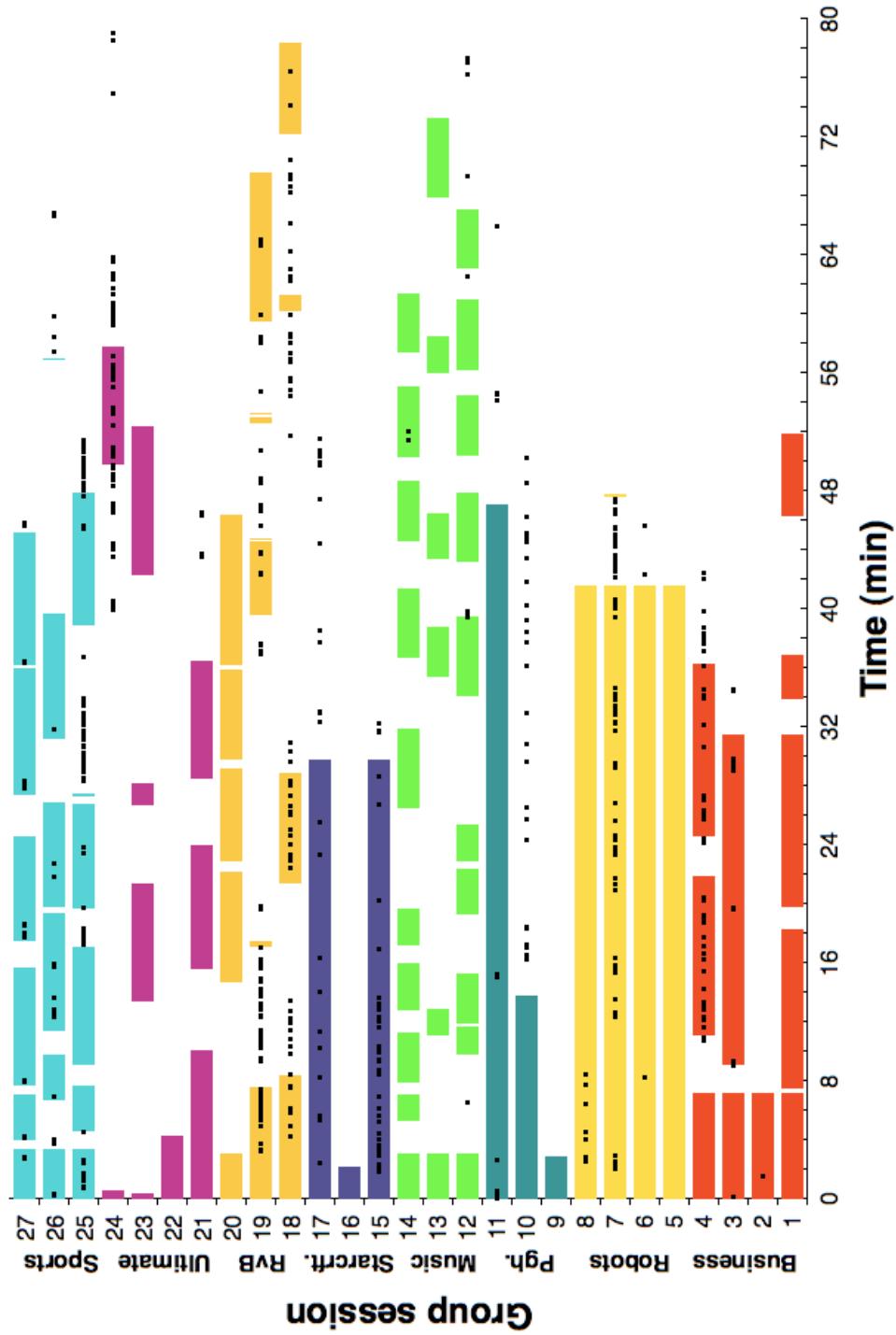


Figure 12-6. Patterns of watching and chatting in the HIT sessions. Colored bands indicate video, gaps between bands indicate waiting, and black dots indicate chat. Only the first 80 minutes of activity are shown.

normalized to the point at which the first video started playing in each group.

There are a few consistent patterns of how participants in each session spent their time between watching and chatting. These patterns are summarized in Table 12-2. In a quarter of the sessions, participants chatted at all points in time: before, during, between, and after the videos. In another quarter, participants generally chatted before or between videos, but did not chat much during them. Participants in eight sessions did not chat, although comments from participants suggest that they may have chatted out-of-band (e.g., using instant messaging) or by talking while physically co-located.

Table 12-2. Summary of watching and chatting patterns.

Pattern	Sessions	%
No chat (or possibly out-of-band)	1, 5, 9, 13, 16, 20, 22, 23	30
Chat before, during, between, and after videos	4, 7, 8, 15, 17, 18, 24	26
Chat mostly before, between, or after videos; little chat during videos	3, 10, 12, 19, 25, 26, 27	25
Not much chat or unclear pattern	2, 6, 11, 14, 21	19

Many groups did not watch the videos immediately one after another. This observation is clearly seen in sessions 12, 13, and 14, and 25, 26, and 27. Participants in these sessions tended to wait varying amounts of time between each video. Some participants may have experienced technical difficulties in watching videos, as shown by the thin vertical strips in sessions 19 and 25.

Figure 12-7 shows a comparison of how much chat occurred while individual participants were playing, waiting, or paused (as the pause feature only paused an individual, and not the whole group). This figure reveals three types of groups: those with members that chatted both while playing and waiting (e.g., ranks 1-6), those with members that primarily chatted during the videos (e.g., ranks 7-9), and those with members that did not chat or possibly chatted out-of-band (ranks 21-27). Only five groups had members that chatted at all while paused, although these group members did not chat much during the pause periods. These results suggest that group members did not use the pause feature to create opportunities to chat without being distracted by the video. One explanation is that participants were simply unaware of the pause feature, although this hypothesis is not supported by

the data. Participants in 21 of the 27 sessions paused the video at least once, and they paused an average of 4.0 times (SD = 4.9 pauses) in each session.

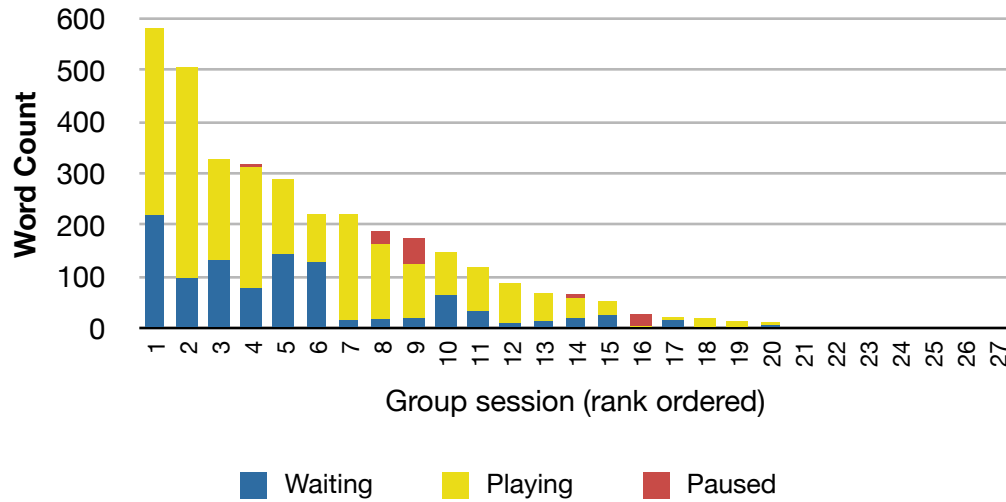


Figure 12-7. Amount of chat while watching the videos, paused during a video, or waiting to begin the next video.

Enjoyment. Overall, participants enjoyed watching the videos and chatting with their friends (M [SD] = 4.21 [.77] of 5). Participants rated the videos they watched positively (M [SD] = 3.82 [.75] stars). Entertainment videos were enjoyed more (M [SD] = 4.1 [.64] stars) than informational videos (M [SD] = 3.42 [.76] stars), $F(1,17.22) = 4.48, p < .05$.

We use a linear regression to understand which factors predict a group’s enjoyment. The explanatory variables were, for each group: the average of their video ratings, the number of videos they watched, a categorical variable coding for whether they watched entertainment or informative videos, a log-transformed count of the number of words spoken in chat, and a control variable for group ID, as discussed earlier. This model has an $R^2 = .55$ (R^2 adjusted = .46). The average rating of the videos watched was a significant predictor of enjoyment, $F(1,16.7) = 9.4, p = .007$. The amount participants chatted with each other, the type of videos watched, and the number of videos watched were not significant predictors.

12.6.6. DISCUSSION

This study shows that many groups did take breaks between videos when they were in control of their own video playback (RQ 12-2), although they rarely took breaks

during videos by pausing (RQ 12-3). Participants in roughly 25% of the sessions chatted throughout the entire experience – before, during, between, and after the videos. Another 25% chatted mostly before, between, or after the videos, with very little of their chat during the videos. About 20% of groups did not use the chat feature enough to determine a clear pattern. Finally, 30% of groups did not use the chat feature at all, although based on comments from participants, they may have chatted out-of-band (such as using IM or Skype for voice chat), or talked out loud with each other as they watched the videos in person. Despite these participants, we do see that some participants used the chat feature throughout, and some reserved their chat for the break periods (RQ 12-1). These results support the use of a waiting screen feature as a method for giving viewers opportunities to chat without being distracted by a video. Interestingly, the initial motivation to include the waiting screen was for mitigating the effects of having viewers who were slightly out-of-sync with each other because of buffering or pausing. From these results, we see that the waiting screen was used the same way the intermission periods were used in the Cartoon study (Chapter 8).

Participants’ enjoyment in this study was solely determined by their ratings of the videos they watched. The amount they chatted, the types of videos they watched (informational vs. entertainment), and the number of videos they watched were not significant predictors of their enjoyment. Thus, it seems that viewers’ enjoyment was solely determined by the subjective quality of the videos they watched, and not any characteristics of their social experience. However, this study is somewhat limited in generalizability because of the small number of groups. In addition, this study’s real-world nature prevented us from controlling potential confounding variables, such as chatting out-of-band. Therefore, this observation needs to be followed up with future research. It could mean that people stopped chatting when they were engaged in watching (as evidenced by the groups that chatted outside of, but not during, the videos), or simply that other attributes of the chat (such as the specific topics discussed) are better measures of the quality of the experience.

12.7. INTERFACE REDESIGN

After running the user study discussed in the previous section, we received feedback from our participants about aspects of the system that were confusing or didn’t work for them. In general, the size constraint placed on our interface from being

embedded in the Facebook canvas page made it difficult for users to find their friends, use the activity visualizations, and find videos to watch. In the initial implementation, these activities were generally exclusive because they required users to refresh web pages and open new web pages to find videos on YouTube.

Once the user study was completed, we set our efforts on redesigning the interface to take advantage of a larger browser window. This redesign was made possible by the release of the Facebook Connect API for Flex applications, an event that occurred after we had already begun implementing the initial interface. Using this API, the new design offers more features directly in the interface, such as seeing friends’ status and finding videos. Figure 12-8 (top) shows an early mockup of the new interface, and Figure 12-8 (bottom) shows the current prototype implementation.

By virtue of not being embedded on the Facebook canvas page – with its toolbars and advertisements – this interface takes advantage of the additional width to display a larger video. It also swaps the positions of the chat box and the playlist to make each larger, and prominently displays the “Quick Add” URL input field so that adding videos is easy.

This interface also adds a tab bar at the bottom (expandable by clicking the disclosure arrow) to display two groups of information. The first tab group contains features that make it easier for users to invite their friends into the group, join the groups their friends are in, and find videos to watch. Mockups for these features are shown in Figure 12-9. The second tab group contains features that help users visualize other users’ activity. This group contains space for displaying information about the current video. It also contains space for the audience representation and chat summarization features discussed in Chapter 11: a scrolling list of recent chat messages, a tag cloud of popular chat topics, and the audience proxy. We decided to make the bottom tabs disclosable to help users manage their attention while watching a video. For users who want to immerse themselves in the video (or their chat group), they can hide the visualizations of other users in the system. For users interested in monitoring the activities of other users in the system, they can make the visualizations visible.

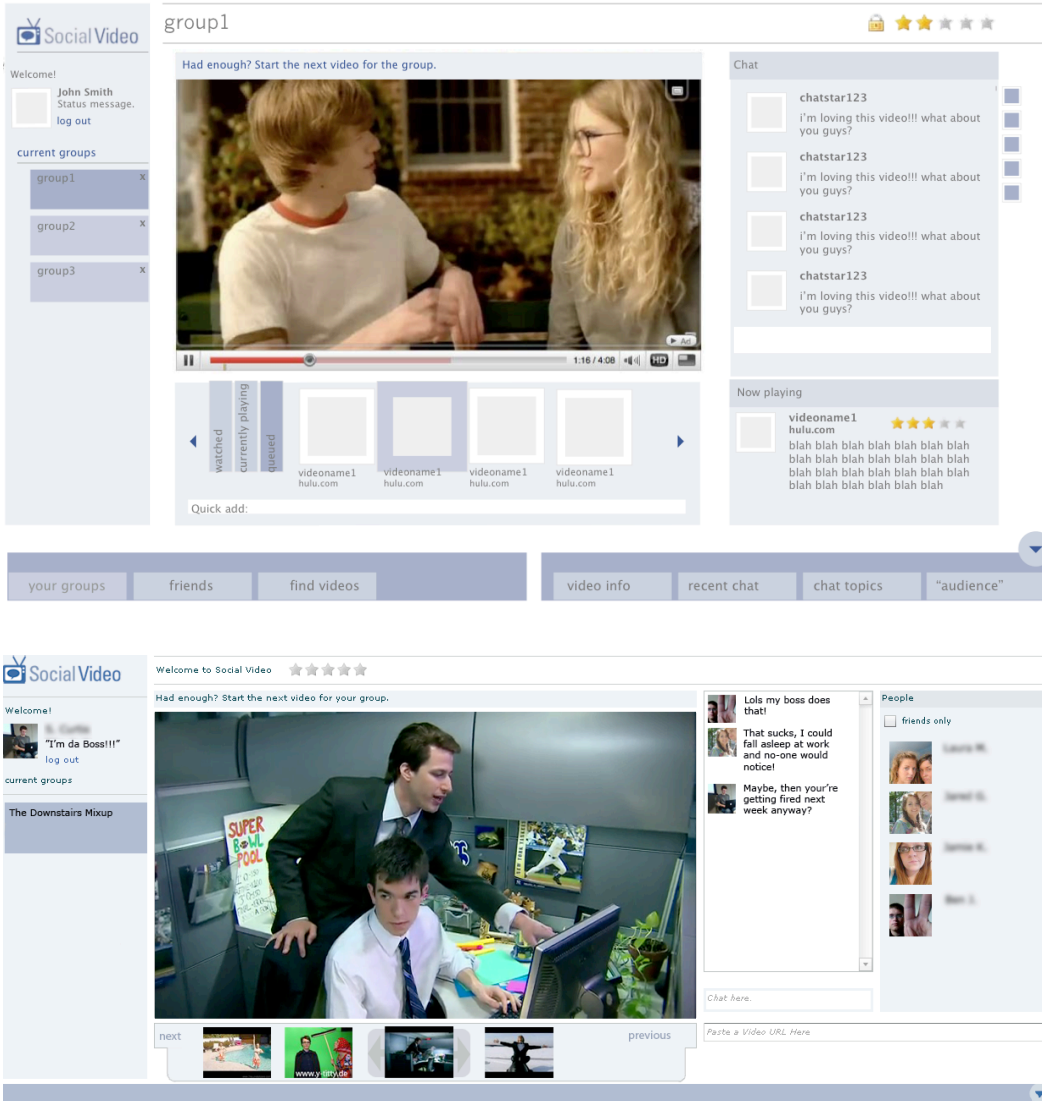


Figure 12-8. (top) Mockup of the redesigned Social Video interface. Tabs are used on the bottom to provide access to friends, videos, and visualizations of other users' activity. The blue boxes on the right are a placeholder for a list of people in the current chat group. (bottom) Screenshot from the current implementation of the redesigned Social Video interface. Names have been blurred for anonymity.

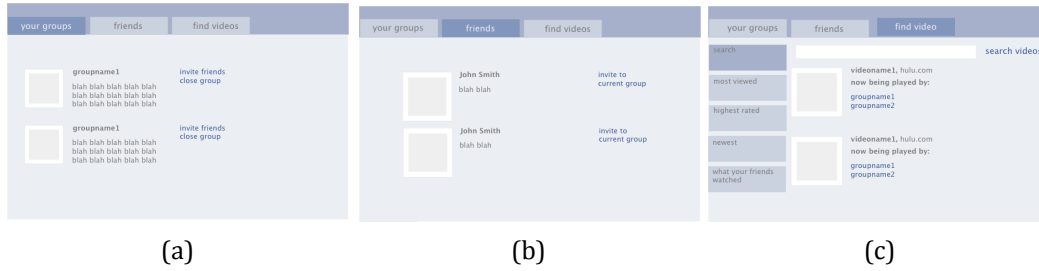


Figure 12-9. Prototypes of the (a) groups, (b) friends, and (c) videos tabs. These tabs are displayed below the video and can be used without interrupting video playback or triggering a page refresh (unlike the initial interface).

As of this writing, the implementation of this interface is ongoing and thus has not been deployed. Despite this fact, the think-aloud usability study described in Chapter 11 validated the ideas of the audience social proxy representation and the visual chat summaries. As discussed earlier, those summaries need not be mutually exclusive. The use of tabs in this interface allows viewers to access both types of summaries (and potentially more in the future). In addition, the design recommendations discussed in Section 11.5.3 – making the audience representation and summaries interactive – are easy to adopt in this interface. Thus, this new interface satisfies the original design goals, summarized in Table 12-3.

Table 12-3. Social Video design goals and how they are achieved in the redesigned user interface.

Goal	How achieved
Goal 1: Provide a fun and sociable experience for all online video	Design supports videos from any online video site that allows embedding; implementation supports YouTube videos. Text chat feature used to promote fun and sociability for viewers watching together.
Goal 2: Support watching with friends	Integration with Facebook provides access to network of friends; the friends list (initial design) and friends tab (redesign) make it easy to invite friends into groups.
Goal 3: Encourage encounters with strangers	Tag clouds and scrolling lists summarize chat activity across the system; Linkages between chat summaries and groups enable viewers to chat with strangers if they are so inclined.
Goal 4: Incorporate large audience features	Redesign includes a tab for the large audience social proxy discussed in Chapter 11.
Goal 5: Maintain ecological validity	Viewers watch videos from YouTube; viewers use real-time text chat common to sites like UStream.TV and Justin.TV.

Goal	How achieved
Goal 6: Implement as few control/authority mechanisms as possible	Viewers successfully used the waiting screen to coordinate their video playback. Whether authority mechanisms are needed to curb undesirable behaviors is unclear, as such behaviors were not observed.

12.8. GENERAL DISCUSSION

This chapter presents the design “Social Video,” a collaborative online video watching application. Its primary goals were to support watching videos with friends, to promote encounters with strangers, and to provide awareness of the community-at-large when there are potentially millions of viewers.

Overall, the design of Social Video met the needs of the users discussed in the scenarios at the beginning of the chapter. By integrating with Facebook, the application was able to tap into peoples’ real-world social networks and make it easy for friends to watch each other. Synchronous chat was used to promote interaction among viewers in the space of a people-centric chat group, instead of an asynchronous, video-centric commenting feature. Synchronous interactions have a positive impact on peoples’ feelings of attachment (Slater, Sadagic, & Schroeder, 2000) and closeness to each other (Chapter 8). Many participants in the studies discussed in Part II enjoyed the combination of synchronous chat while watching videos. Thus, this design decision supported the goal of providing a fun and sociable experience for users through the activity of chatting while watching videos (Goal 1).

Although relative use of the application was low, we were able to learn quite a bit from the user study. The fact that participants used the waiting period in the same way as the intermission periods in the Cartoon study provides strong support for the inclusion of this type of feature for collaborative online video systems. The fact that participants generally used these intermission periods to chat, rather than creating their own periods by pausing, suggests that the intermission periods ought to be scheduled by a moderator or community leader, rather than the viewers themselves. This notion is akin to the structure of Tutored Video Instruction (Gibbons, Kincheloe, & Down, 1977), in which teachers are responsible for pausing a video at a specified time to give students an opportunity to discuss. In the online realm,

distance learning applications may still require the interventions of a teacher to engage students in a virtual classroom in a discussion.

Developing applications for a real-world audience is a difficult task, and small problems such as technical difficulties or user interface glitches can easily discourage potential users. We were fortunate to be able to attract several early adopters from our study with Mechanical Turk users (indeed, several continued to use the application even after the Turk study had ended). Their feedback provided us with many ideas for how to improve the application. This feedback, coupled with additional usability testing, resulted in the redesigned interface discussed in Section 12.7. Although a completed implementation, deployment, and evaluation of this interface remains as future work, it is my hope that the ideas about helping people navigate in and maintain awareness of a large audience make their way into the designs of future collaborative online video applications.

12.9. SUMMARY AND CONCLUSIONS

- This chapter presents the design of “Social Video,” a collaborative online video watching application. Watching and chatting among friends is supported by integrating with their existing social networks. Interactions with strangers are encouraged by making users’ activity visible in visualizations and by allowing users to join multiple chat groups simultaneously.
- To promote back-and-forth discussion, the application uses a group-centric approach in which viewers watch videos and chat in groups using a text chat feature.
- To promote awareness of the other users in the system, the application implements the tag cloud and scrolling list chat summaries discussed in Chapter 11.
- In an evaluation study, groups of friends watched a series of videos with each other. They used time between videos to chat without being distracted by the videos. They did not pause to create break periods during videos. These findings provide real-world support in favor of the use of intermission periods to provide opportunities for viewers to chat without being distracted by the videos.

- Participants' enjoyment was significantly predicted by their ratings of the video content. Although the amount participants chatted was not a factor in their enjoyment, several groups may have chatted out-of-band, masking a potential effect.

Part IV: Learning About Videos

Chatting while watching videos can be construed as a human computation process. The activity of chatting while watching exhibits an algorithmic behavior: for the input of a video, people who watch that video collaboratively produce an output in the form of textual chat transcripts that can be used to learn meaningful information about the video, such as tags, ratings, and profiles of moment-by-moment enjoyment.

13.

LEARNING ABOUT VIDEOS FROM CHAT DATA

One central tenet of this dissertation is that useful information about videos can be mined from the chats of people who have watched those videos with others. When viewers chat with each other while watching a video, their utterances produce a corpus that can be analyzed for information about the video. As we saw in the *Cartoon and Text vs. Audio* studies, about 36-40% of chat was focused on the video being watched. Thus, because viewers are talking about the video, their language – the words and phrases they use in their chat – may be descriptive of the video.

The chapters in Part IV address the overall question of can be learned about a video from chat data. These chapters demonstrate that three items that can be learned: a set of tags that describe a video (Chapter 14), hints about viewers' enjoyment of a video that can be used to infer video ratings (Chapter 15), and profiles that show which parts of a video viewers most enjoyed (Chapter 16).

Learning about videos through chat has several benefits for online video communities. Foremost is that many online communities suffer from problems of under-contribution (Butler, 1999). Online communities depend on contributions from their members in order to thrive. Without new contributions, an online community's content remains static. Without new members making contributions, an online community's population will dissipate. Thus, increasing both the number of contributions made, and the number of contributors, is an important goal for all online communities (Butler, 1999; Ludford et al., 2004; Cosley et al., 2005; Harper, Sen, & Frankowski, 2007; Lee et al., 2009).

Methods that automatically infer information about videos can be used to increase both of these quantities. Learning about videos through chat data transforms an explicit contribution (e.g., by typing in a tag or rating a video) into an implicit one. No action is required of viewers beyond that in which they already engage while watching videos collaboratively. By increasing the scope of the definition of a “contribution,” we can increase both the number of people who contribute, and the number of contributions they make. Thus, the ability to receive implicit contributions is a desirable property for online video communities.

The only exception to this process is when viewers do not utilize the chat feature. This exception may happen because of non-interest (discussed in Chapter 6) or distraction (discussed in Chapters 7-9). In cases of non-use, no information can be inferred about the viewers’ enjoyment or perceptions of the video. However, for those viewers who do engage in chat, they need not be further distracted or bothered by the need to label or provide metadata about the videos they watch; this information can be measured unobtrusively, directly from their chat.

Unobtrusive measures are desirable because they avoid a self-report bias (Cozby, 2004). These biases can occur for a variety of reasons. For example, when rating an item, people may be uncertain or unaware of how they feel toward the item, uncertain of how to evaluate the levels on the rating scale, indifferent to providing an accurate rating, or purposefully inaccurate to avoid a socially-undesirable answer³⁰. Unobtrusive measures avoid these biases because they measure a person’s behavior instead of their attitudes. For example, although a person watching a movie may claim she is not enjoying it, the smile on her face belies her true feelings. Unobtrusive measures are desirable in an online setting because the cost of providing inaccurate information or not providing any information at all is nil, and the ability for others to detect inaccuracies is limited. In addition, the quality of information obtained unobtrusively may be higher. In Chapter 14, I will discuss a situation in which implicit contributions through chat results in a set of tags that are generally of a higher quality than those that were entered manually.

One final benefit to learning about videos through chat data is that it can make it possible to infer information about a video that is costly to collect directly. Many television and movie producers screen their videos in front of test audiences to learn

³⁰ This case may occur when, for example, one does not want one’s friends to know that he or she watched, or enjoyed watching a particular video.

about their moment-by-moment enjoyment of the film or show (Eliashberg & Sawhney, 1994). This information is beneficial because it feeds back into the creative process; if a scene is unappealing, it can be improved or eliminated in the final cut. Collaborative online video watching opens this application to a wider Internet audience. In Chapter 15, I discuss how chat data can be used to create enjoyment profiles of videos based on aggregated laughter across viewers. These profiles depict moment-by-moment levels of enjoyment of a video, which can be used to improve the quality of video search and recommendation engines and their user interfaces.

13.1. ONLINE SOCIAL INTERACTIONS AS HUMAN COMPUTATIONS

The general method by which the information described above is learned – tags, ratings, and enjoyment profiles – is through a human computation process. Human computation is a field of computer science devoted to figuring out how to insert humans into a computation process to augment or collect data for machine learning algorithms. In this case, humans act as a supplemental intelligence for the algorithm by performing steps that would otherwise be difficult or impossible for a computer, or would produce inaccurate results. This supplement often feeds back into the improvement of machine learning algorithms by providing hand-labelled training examples. Thus, these algorithms improve their performance by having humans demonstrate correct responses.

The notion of having humans perform computationally-intractable tasks is often encapsulated in a CAPTCHA task (von Ahn, Blum, & Langford, 2004a). People are motivated to perform this task because the task serves as a gatekeeper to a desirable resource, such as free email or pornography. Other human computation tasks involve playing games with others, or against a simulated computer, to collect information about textual, visual, or auditory inputs. These games have been used to collect a wide variety of data, including the following below.

- ESP Game: labels for images (von Ahn & Dabbish, 2004b)
- Peekaboom: locations and labels of objects within an image (von Ahn, Liu, & Blum, 2006c)
- Phetch: textual descriptions of images (von Ahn et al., 2006a)
- Verbosity: common-sense facts (von Ahn, Kedia, & Blum, 2006b)

- Search War: relevance judgements for search results (Law, von Ahn, & Mitchell, 2009b)
- TagATune: tags for music clips (Law & von Ahn, 2009a)

In the next chapters, we will see a new context for human computation: that of online social interaction. The chat messages exchanged between participants in an online conversational space are shown to be a new source of information for learning about items of interest. In this dissertation, those items are videos; however, the methods described can be applied to any domain in which people can be motivated to converse about the items.

13.2. COLLECTING AND MINING CHAT TRANSCRIPTS

The process of learning about a video from chat is depicted in Figure 13-1. This process possesses an algorithmic behavior. Inputs to the algorithm are people and a video, and outputs are a sets of tags, ratings, and profile of enjoyment. In the intermediate stage, the activity of collaborative watching is used to produce chat transcripts for the video. Although the chat transcripts need to be textual in nature, this requirement does not restrict the media used for the actual chats. For example, participants could communicate using a voice chat feature while watching a video, and the system could use speech-to-text technology to create a textual transcript. If this transcript were created in real time, chatters could be enticed to correct transcription errors themselves³¹.

After the chat transcripts have been gathered for a video, they can be analyzed and mined for information of interest. Chapter 14 discusses how to use term weighting metrics to extract a set of relevant tags for a video. Chapter 15 discusses the use of linguistic analysis to infer a rating for a video. Chapter 16 demonstrates that aggregated laughter in chat can be used to build profiles of video enjoyment over time. Each of these chapters demonstrates that useful information about videos can be learned directly from chat data, in an implicit fashion.

³¹ This correction could be part of a separate human computation game for improving the quality of speech-to-text translation systems. Providing incentive to perform corrections, as well as creating an interface that does not significantly add distraction to the video-watching experience, are both important design challenges.

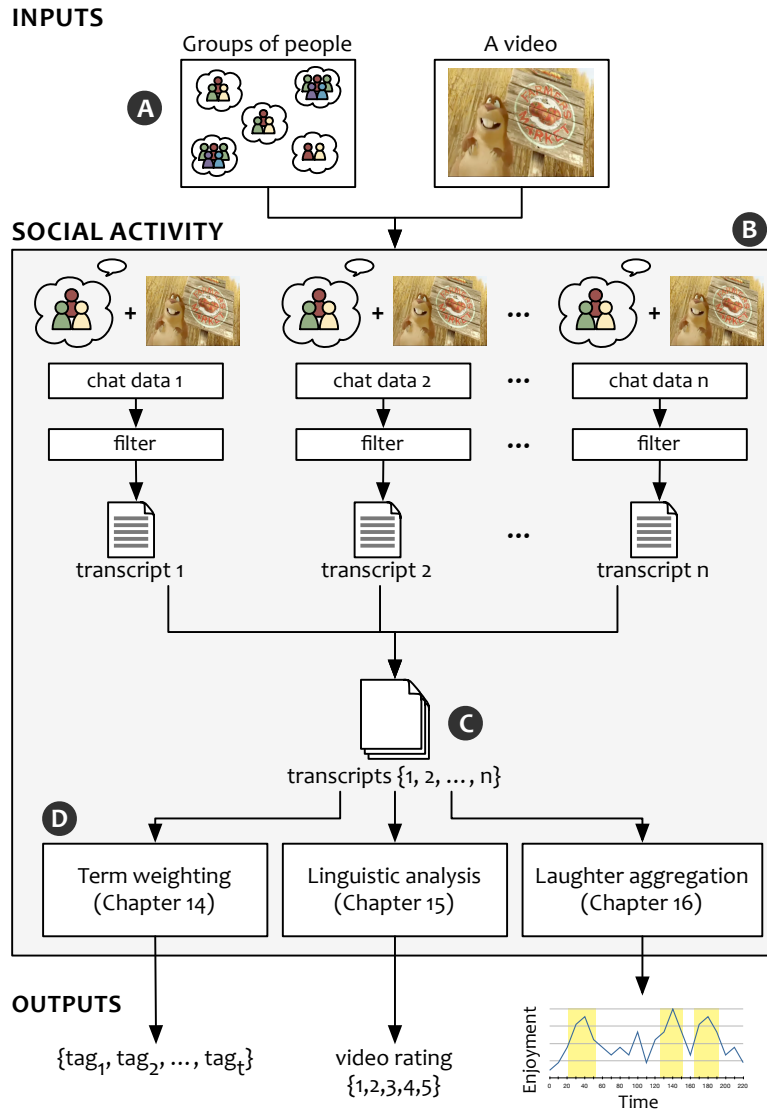


Figure 13-1. Algorithmic process for collaborative online video watching as a human computation. (a) Inputs to the algorithm are groups of people and a video. (b) The social activity of watching videos and chatting is used to produce a chat transcript for each group. A chat filter can be used to convert voice chat into a text transcript, and text transcripts can be normalized before analysis. (c) Chat transcripts are a by-product of the social experience. (d) Various analyses are performed on the transcripts to produce the outputs of tags, ratings, and enjoyment profiles.

13.3. SUMMARY AND CONCLUSIONS

- Human computation processes are used to supplement machine learning algorithms by having people generate data for or perform the computations of tasks that are difficult or impossible for computers to perform.
- Examples of human computation processes include learning labels for images, identifying objects in images, tagging musical clips, and rating the relevance of search results.
- Human computation processes are often instantiated as games in which one or more players, through their actions in the game, provide machine learning algorithms with training data. These games often require that players do not communicate with each other, eliminating their ability to foster social interactions among players.
- This chapter contends that human computation processes can be carried out in the presence of social interaction. In the domain of collaborative online video, groups of people and a video are inputs to the computation. As viewers watch the video and chat with each other, they produce a set of chat transcripts. These transcripts can be mined for information about the videos, including a set of tags (Chapter 14), ratings (Chapter 15), and profiles of moment-by-moment enjoyment over the course of the video (Chapter 16).
- Learning information about videos implicitly, from chat transcript data, is one way that online video communities can increase the number of contributions they receive (i.e., information about their videos), and increase the number of members who make contributions.

14.

TAG EXTRACTION

Tags are short words or phrases used to describe an item. Tags are used in many online domains, including: videos (YouTube), email (GMail), bookmark management (Delicious), photo sharing (Flickr, Facebook), blogging (Technorati), and news and link aggregation (Slashdot, reddit, Digg). Tags help people understand the items they are browsing and locate other related items. Tagging is different from traditional information classification methods, such as hierarchical classification, as tags are applied directly by users, often without any guidelines, rules, or restrictions for what constitutes a valid tag (Macgregor & McCulloch, 2006). Despite the free-form nature of tagging, a number of researchers, bloggers, and technology critics have explored questions of how tagging systems compare to expert-constructed ontologies (Shirky, 2005), what kinds of tags users apply (Sen et al., 2006; Heckner et al., 2008), and whether tagging systems help users find information (Furnas et al., 2006; Chi & Mytkowicz, 2008). The overall conclusions of their work are that tagging is here to stay, and numerous challenges remain in eliciting relevant and efficient tags from community members. Challenges are also present in selecting which tags to display when there are many choices (Sen et al., 2007).

This chapter discusses a method for automatically tagging videos in a collaborative online video site. This method extracts tags from viewers' chat data by searching for the most popularly and uniquely used terms in the chat corpus. In my evaluation of this method, I compared the quality of chat-extracted tags to tags hand-applied by people. I found that tags extracted from chat are generally as good as those applied by video uploaders on YouTube. Thus, automatic tag extraction behaves as a human computation process: a set of tags can be inferred for a video from the conversations viewers have with each other as they watch together.

14.1. GENERAL METHOD FOR TAG EXTRACTION

Extracting tags from a corpus of chat transcripts is akin to a summarization task: which subset of words or phrases best describes the subject matter? One common method used by summarization technologies is to segment the text in the corpus in some meaningful way (e.g., by sentences or paragraphs), compute a metric of importance for each segment, and then select the highest ranked segments to build the summary (e.g., Nenkova, Vanderwende, & McKeown, 2006).

In the task of extracting tags, we are interested in figuring out which words or short phrases best describe the video. Thus, our segments include both unigrams (single words) and bigrams (word pairs). For clarity, I refer to both unigrams and bigrams as "terms" in the corpus. The task is now to figure out which terms best describe a video by using a metric to rank the terms in the corpus of chat transcripts. Figure 14-1 depicts this task graphically. The general strategy is to pick terms that are frequently used within one video and not frequently used across other videos.

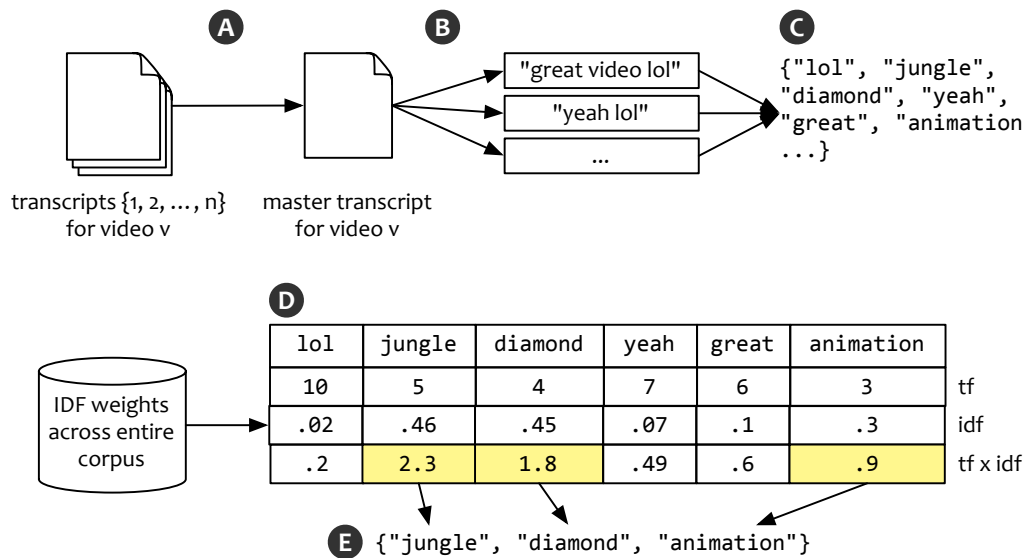


Figure 14-1. Process diagram for extracting tags for a video from chat transcripts. (a) Multiple chat transcripts for a video are combined into a single transcript. (b) Individual comments are normalized. (c) Unigrams and bigrams are extracted from the comment set. (d) Each term is weighted; in this example, the TF-IDF metric is used to weight terms. (e) The top K terms are chosen as the extracted tags.

The inverse document frequency (IDF) metric is commonly used in computational linguistics research as a measure of the value of, or content contained in, a word in a corpus. The intuition behind this metric is that words that occur in relatively few

documents (i.e., chat transcripts for a video) represent the unique content communicated by that document, and therefore what makes that document have value. Words that occur in many documents have lower IDF weights than infrequently occurring words. For example, the word "yeah" is very common in the English language, and thus would have a low IDF weight. By contrast, the word "jungle" is less common and would have a higher IDF weight.

One variant on the IDF metric is the TF-IDF metric (term-frequency IDF). This metric weights IDF scores by the frequency of a term's occurrence. Thus, terms that occur more frequently are weighted higher than terms that only occur once or twice. We use this metric for weighting terms to account for the fact that terms are often repeated within and across chat messages. For example, consider the message: "He has this big head little legs and little arms and trying to fight everthing." (sic, participant C44, Cartoon study). In this message, the term "little" is given more weight because it is mentioned twice. IDF alone would not account for the second mention of "little."

We use a "bucket of words" model for representing the terms spoken in chat. This model is not concerned with the ordering of messages, the ordering of words within messages, the particular speakers of messages, or the groups in which messages were spoken. Rather, chat messages spoken during each video are combined and treated as the "documents" in the corpus. This strategy makes it easy to find terms that are frequently used within one video and infrequently used across other videos. Thus, a single IDF weight is computed for each term in the corpus, and TF-IDF weights are computed for each term used in each video in the corpus. Formulas for these computations are given in Figure 14-2.

$$\begin{aligned}
 V &= \{\text{all videos}\} \\
 n_{t,v} &= \text{count}(\text{occurrences of term } t \in \text{video } v) \\
 TF(t, v) &= \frac{n_{t,v}}{\sum_{t' \in v} n_{t',v}} \\
 IDF(t) &= \log \frac{|V|}{|\{v \in V : n_{t,v} > 0\}|} \\
 TFIDF(t, v) &= TF(t, v) \cdot IDF(t)
 \end{aligned}$$

Figure 14-2. Formulas for computing TF-IDF weights. TF (term frequency) is the frequency of occurrence of term t during video v . IDF (inverse document frequency) is the measure of a term's "uniqueness," and is inversely proportional to the number of videos in which that term was used.

14.1.1. TEXT NORMALIZATION

Before computing TF-IDF weights for the terms in each video, several text normalization procedures are performed to increase the likelihood of extracting meaningful terms. First, common stop words and punctuation symbols are removed from the corpus. Stop words are contentless function words, such as "the" and "it," and are often dropped prior to substantive text mining procedures. The remaining words were stemmed using the Porter stemming algorithm (Porter, 1997). This process removes inflectional endings, such as markers of plurality and tense, in order to generalize across alternative representations of the same word. For example, "swimming" is stemmed as "swim". Finally, bigrams were created for each chat message to handle cases in which short phrases were more commonly used than their singular counterparts (e.g., "general motors" has a different meaning as a tag than "general" or "motors").

14.1.2. LIMITATIONS

One limitation to this method of extracting tags from chat is the potential for abuse and/or gaming of the system. Because we use TF-IDF as the term-weighting metric, a potentially mischievous user Michael can artificially inflate the uniqueness of an irrelevant term, like "badtag." To do this, Michael can spam his chat room with messages of the form "badtag badtag badtag badtag badtag badtag" to increase the frequency that "badtag" occurs. He can also make "badtag" unique over the entire chat corpus by adding random characters to the end, such as "badtag_123!#". TF-IDF would rank this term highly because of its frequent occurrence in Michael's chat messages, and its lack of occurrence across other chat messages.

Several strategies can be used to combat this problem. First, text processing methods stricter than those described can be used to negate the effect of adding random characters to a term. For example, punctuation characters can be converted to spaces or completely eliminated, transforming "badtag_123!" back to "badtag." However, this step does not completely eliminate Michael's ability to make "badtag" unique in the corpus; he could simply append a string of random letters to the end, as in "badtagaaaaaaa." In this case, the use of a tag white-list (e.g., a dictionary with pre-defined acceptable tags) would prevent Michael from introducing his "badtag" into the system. Since a dictionary does restrict the possible tags that could be applied to a video, its use should be informed by the possible risk or concern for a

malicious tagger. Other strategies can also be used to restrict or eliminate malicious tagging. These include using CAPTCHAs to prevent users from creating many malicious accounts, banning users, using user-level weights on tags, or even manually approving tags before they are applied to a video.

14.2. TAG EVALUATION STUDY

The previous section describes a process for extracting a set of tags that describe a video from raw chat log data. But, do the tags extracted using this method actually describe the video, or are they irrelevant?

To answer this question, human raters were recruited to participate in the Tag Evaluation study. In this study, they watched videos and rated the relevancy of a set of tags to the video. Relevancy was defined as how well each tag fit the video. No experimental conditions were used in this study as its purpose was to have participants simply evaluate the relevancy of tags.

14.2.1. TAG EXTRACTION

The tag extraction process described above was run on the chat data collected for the 15 videos used in the Cartoon and the Text vs. Audio studies; these videos are described in Chapter 5. To understand the quality of the tags extracted from chat, we collected tags for these videos from two human sources: tags applied on the videos' YouTube pages, and tags applied as part of the Best Part Labeling lab study (discussed in Chapter 16). As the purpose of the Tag Evaluation study was to collect ratings of the relevancy of tags, we added another source of tags as a manipulation check: red-herring tags that were applied to *other* YouTube videos. By design, these tags were irrelevant to the videos to which they were applied.

Two challenges were present when creating tag sets for comparison: not having enough data and having too much data. The former case occurred only when collecting tags from YouTube. Five of the videos did not have YouTube pages³², and thus tags from YouTube could not be collected; these videos were dropped from the study, leaving us with a set of 10 videos to use in the study. How the latter case was

³² Note that although the Cartoon study videos originally came from Channel Frederator, many of them had since been uploaded to YouTube when this study was conducted.

handled depended on from where the tags came. Tags that were extracted from chat and tags that were applied by hand had the benefit of being weighted by the TF-IDF metric or frequency of occurrence, respectively. In these cases, the five tags with the highest weights were chosen for the evaluation set. Tags from YouTube did not come with any measures of importance or weight; thus, five tags were chosen at random to constitute the evaluation set. Certain tag variations were manually excluded when picking randomly to avoid repetitious tags (e.g., once “stopmotion” was chosen, “stop-motion” was excluded from being chosen).

In many cases, identical tags were selected in multiple sources. In the analysis, we treated those tags as coming from a separate, common source. Common-source tags were normalized with respect to capitalization to make a fair comparison. Thus, a tag of “pen” from YouTube and a tag of “Pen” from chat were treated as the same tag.

The tag selection process resulted in five sets of tags for each video:

- Common tags that occurred in multiple sources,
- Chat tags extracted from the chat transcripts,
- Human tags applied by participants in the Best Part Labeling study (Chapter 16),
- YouTube tags applied by video uploaders on YouTube,
- Herring tags applied to unrelated videos on YouTube.

Note that two of these sources – Human and YouTube – represent cases in which people explicitly perform the tagging operation. Human tags were applied by people in the context of a lab study, and YouTube tags were applied by people on the open Internet.

Table 14-1 details all of the tags selected from each source for each video in this study. Note that the tags are shown in their original form, with capitalization and spelling unaltered. An artifact of this preservation is that some tags are plurals of others (e.g., “daughter” and “daughters”). These cases were preserved as relevance of the singular form may not imply relevance of the plural form. There were no cases in which two tags differed only in their capitalization.

Table 14-1. Tag sets for each video in the Tag Evaluation study. Common tags marked with an asterisk (*) also belonged to the set of tags extracted from chat. Common tags marked with a dagger (†) also belonged to the set of tags from YouTube. All common tags also belonged to the set of tags from human raters. (‡) This tag was unintentionally relevant to the video and thus was excluded from the analysis.

Video	Tags	Time
Ali G - War	Common: war†, general†, Ali G† Chat: ali, canada, nuke, motor, general motors Human: scowcroft, interview YouTube: motors, stowcraft Herrings: Nursery, debarge, Minister, launched, racing	5:59
Brothas From the Same Motha	Common: martian*†, marvin*†, bic*† Chat: brothas, bic guy Human: cartoon, motha YouTube: thewinekone, brothers Herrings: Handmaid's, vencedores, ame, ariana, Molasses	4:01
Daughters	Common: daughter†, daughters* Chat: 12, daughters right, 12 YEARS, watching daughters Human: condom, comedy, condoms YouTube: terrorist, kidnap, 24, bomb Herrings: diablo, woody, village, multitrack, neoprene	3:35
Fuggy Fuggy	Common: fuggy*†, ninja* Chat: FUGGY FUGGY, f*ck, fuggy means Human: japanese, cartoon, leisure YouTube: star, rest, poo, toilet Herrings: cia, statistics, menu, across, Kearney	4:53
Gopher Broke	Common: gopher*†, animation Chat: gopherbroke, Cow, CHICKENS, tomato Human: funny, cartoon, market YouTube: trailer, anime, film, anon Herrings: observatory, Martyrs, hansen, Isaac, chamber	4:18
In the Rough	Common: animation†, rough† Chat: Sko, Diamond, jungle, claymations, unibrow Human: caveman, cartoon, funny YouTube: in, kids, blur Herrings: fishtank, results, necklace, Instrumetal, mariah	4:49

Video	Tags	Time
Paddy the Pelican	<p>Common: pelican*†, bears*, paddy*†</p> <p>Chat: Popeye, buster</p> <p>Human: cartoon, boat</p> <p>YouTube: wet, cartoons, Sam</p> <p>Herrings: handcuffs, Coasters, Hippo, fossils, Skittles</p>	5:13
Pen Pals	<p>Common: pen*†, pencil*</p> <p>Chat: panni, Sumo, sharpening</p> <p>Human: love, animation, the</p> <p>YouTube: Pals, aniBoom, at, Funny</p> <p>Herrings: off-road, Challenger, DonOmar, unlocking, rangers</p>	5:07
Plumber	<p>Common: plumber*†</p> <p>Chat: ducky, NOSE, fingers, lisa</p> <p>Human: water, leak, plumbing, cartoon</p> <p>YouTube: short, movie, red, pe</p> <p>Herrings: Pandemic, roaming, execution, gracie, mabeline</p>	6:04
Tony vs. Paul	<p>Common: tony*†, paul*</p> <p>Chat: sliding, Grass, Pretty cool</p> <p>Human: fight, motion, friends</p> <p>YouTube: stopmotion, battle, animate, motion</p> <p>Herrings: spoiler, gate‡, pediatric, Handlebars, square</p>	5:02
Total		49:01

Of the 22 tags in the Common category, 15 (30%) were part of the original set of tags extracted from chat and 16 (32%) were part of the original set of tags applied on YouTube. This amount of overlap suggests that many of the words used in chats are relevant to the videos being watched. However, chat extraction does seem to work better for some videos than others. For example, the chat-extracted tags for Gopher Broke (“gopherbroke,” “Cow,” “CHICKENS,” and “tomato”) seem to be more relevant than the chat-extracted tags for Pen Pals (“panni,” “Sumo,” “sharpening”).

14.2.2. PARTICIPANTS

To rate the relevancy of each of these sets of tags to their respective videos, 30 participants were recruited through the CBDR web site. Participants were generally in their mid-twenties (M [SD] = 26.0 [9.6] years). Half of participants were male, and

60% spoke English as their native language. Participants were paid \$15 for their time. Participation in this study took approximately one hour.

14.2.3. METHOD

Participants watched the ten videos listed in Table 14-1 in a room with a projector and speakers. Participants watched the videos in groups of up to six people, and they were instructed to not talk with each other during the videos. The videos were shown in a random order for each group. After each video, participants were given a paper form that asked them to rate the video and the relevancy of the tags for that video. Video ratings were made on a 5-point scale (1-5 stars, 5 highest). Tag ratings were made on a 4-point relevancy scale: “Not relevant at all” (coded as 0), “A little relevant” (coded as 1), “Somewhat relevant” (coded as 2), and “Highly relevant” (coded as 3).

To guard against order effects, three versions of the rating form were created for each video. Each version listed the tags in a different (random) order. As this was a paper-based study, we felt that three permutations of tag orderings provided enough protection against order effects without making the analysis overly difficult and error-prone. Each specific permutation of tags was used ten times.

14.2.4. RESULTS

For each participant and each video, we computed a mean relevance score for each source of tags (Common, Chat, Human, YouTube, Herrings). Tags that occurred in multiple sources (i.e., in the Common category) were only included in the Common category; thus, the mean relevancy scores for Chat, Human and YouTube tags reflect the mean relevancy of only their unique tags. To evaluate the impact of separating out the Common tags, we also computed a mean relevance score for the original set of five tags from each source (AllHuman, AllChat, AllYouTube). This secondary analysis is discussed later in this section.

A analysis of covariance³³ (ANCOVA) is used to compare the mean relevance scores between the different tag sources. The outcome variable for the ANCOVA was the mean per-tag relevance. The explanatory variables were the tag source, the video ID, an indicator variable for whether the participant was a native speaker of English, and all two-way interactions. In addition, the participant ID was added to the model as a random effect to control for within-participant variance. Additional control variables, including a participant's age and gender did not account for significantly more variance in the ANCOVA. Thus, these variables were not included in the final models.

The mean overall relevance for each of the different sources of tags in both models is shown in Table 14-2. The model had an $R^2 = .76$ (R^2 adjusted = .75). A Student's t test showed that each tag source was significantly different from the others.

Table 14-2. Relevance statistics for tag sources. Standard deviations are listed in parentheses. Student's t letters show which tag sources were significantly different at the $p = .05$ level. Tag sources not connected by the same letter were significantly different.

Tag Source	Mean (SD) per-tag relevance	Student's t
Common	2.5 (.65)	A
Human	2.2 (.70)	B
Chat	1.5 (.74)	C
YouTube	1.3 (.87)	D
Herrings	.33 (.43)	E

Common tags were rated the highest, with a mean per-tag relevancy rating of 2.5 (out of 3). Human tags were rated a mean of 2.2 on the relevancy scale. Qualitatively, both of these ratings are between "Somewhat relevant" and "Highly relevant". Chat tags were rated significantly higher ($M = 1.5$) than YouTube Tags ($M = 1.3$), corresponding to between "A little relevant" and "Somewhat relevant". As expected, Herring tags were rated poorly ($M = .33$).

The interaction between being a native speaker and the tag source on mean per-tag relevancy was significant, $F(4,1408) = 8.36$, $p < .0001$. Figure 14-3 shows the

³³ A multivariate analysis of variance (MANOVA) can also be used for this analysis. I use an ANCOVA with stacked data (multiple rows per participant, one for each tag source) as it provides a clearer picture of the contrasts between the different tag sources, albeit with the consequence of inflated degrees of freedom in the F tests.

differences in tag ratings between native and non-native speakers. Contrast testing showed significant differences between native and non-native speakers in their ratings of YouTube tags ($F [1,43.1] = 6.2, p < .02$), Chat tags ($F [1,43.1] = 4.4, p < .05$), Human tags ($F [1,43.1] = 11.5, p < .01$), and Herring tags ($F [1,43.1] = 8.97, p < .01$). The difference between Common tags was not significant ($F [1,43.1] = .06, p = n.s.$).

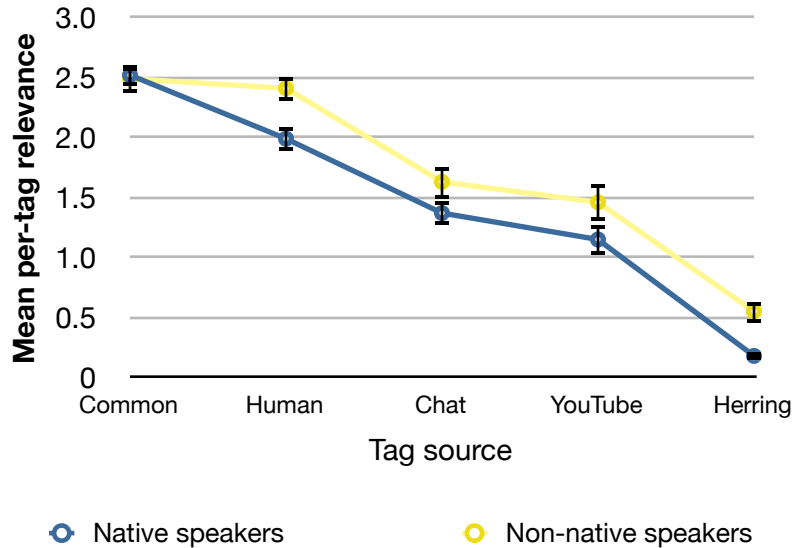


Figure 14-3. Comparison of tag relevancy ratings between native and non-native English speakers. Error bars represent 95% confidence intervals. All differences between native and non-native speakers were significant at the $p = .05$ level, except for Common tags.

The interaction between being a native speaker and the video ID on mean per-tag relevance was not significant, $F (9,1408) = .12, p = n.s.$ Therefore, non-native speakers did not differ from native speakers in their overall relevancy ratings for the tags for each video. The interaction between video ID and tag source was significant, $F (36,1408) = 16.3, p < .0001$. Thus, the relevancy of tags from different sources depended on the video to which the tags were applied. Therefore, we compare the mean relevancies of the tags from each source at a per-video level in Figure 14-4.

Figure 14-4 shows several trends. In general, Common and Human tags were rated the highest. Chat tags were rated significantly higher than YouTube tags for six videos: “Ali G - War,” “Fuggy Fuggy,” “Gopher Broke,” “In the Rough,” “Pen Pals,” and “Plumber.” Chat tags were rated equal to YouTube tags for one video, “Brothas From the Same Motha.” Chat tags were rated significantly lower than YouTube tags for three videos: “Daughters,” “Paddy the Pelican,” and “Tony vs. Paul.” As expected, Herring tags were rated the poorest in every case.

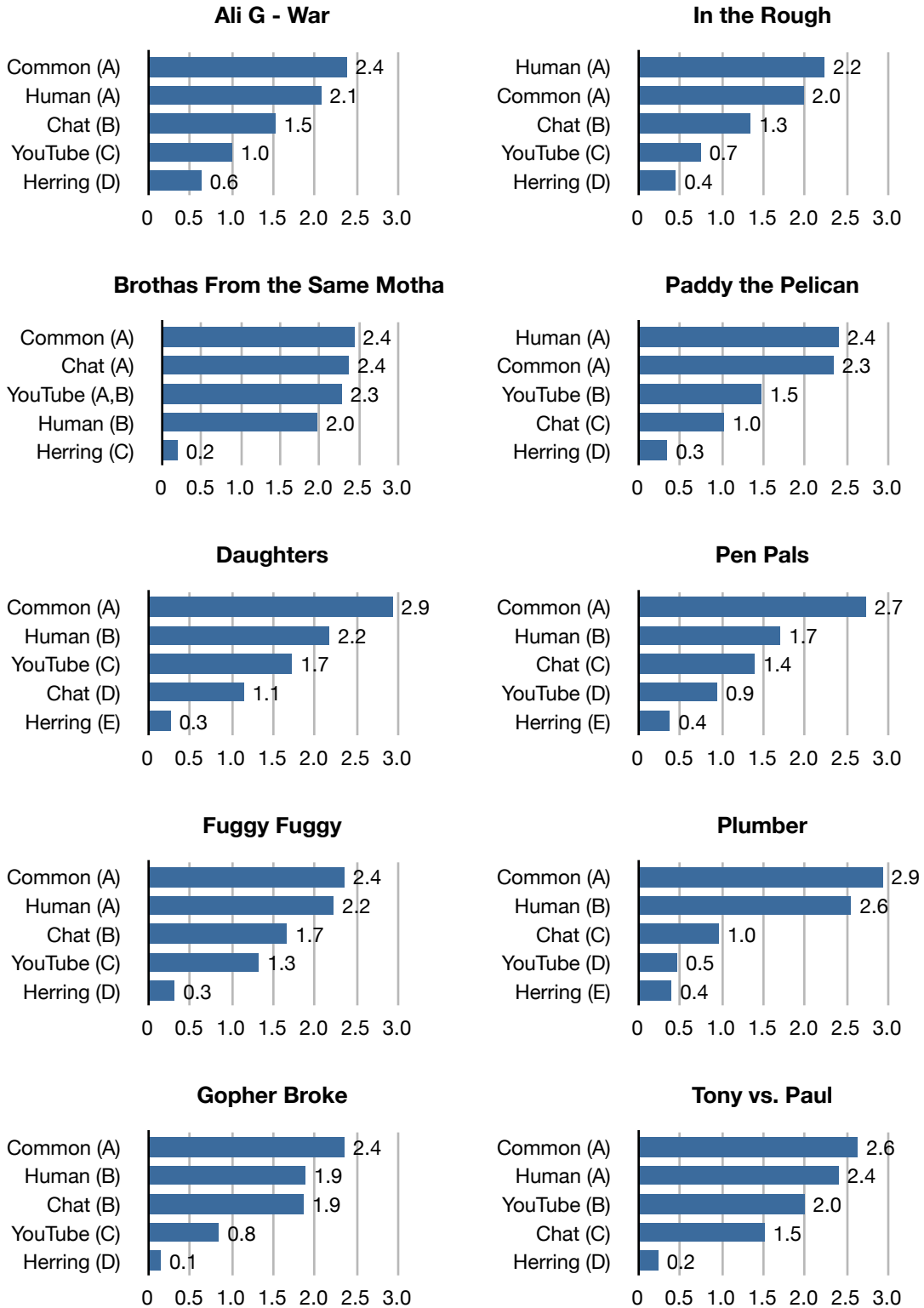


Figure 14-4. Mean per-tag relevance for each tag source and each video. Student's t letters are listed in parentheses and show which tag sources were significantly different at the $p = .05$ level. Tag sources not connected by the same letter were significantly different.

One explanation for why Chat tags were generally rated higher than YouTube tags is that the more relevant YouTube tags may have been part of the Common set and the unique YouTube tags were of a lesser relevance. In other words, the Common tags found on YouTube may have been rated higher than the Common tags that were extracted from chat. To explore this possibility, we first compared the number of tags in the Common set that were also part of the Chat and YouTube sets. This comparison is shown in Table 14-3. This table shows that, across all of the videos, 68% of the tags in the Common set also belonged to the Chat set, and 73% of the tags in the Common set also belonged to the YouTube set.

Table 14-3. Percentage of tags in the Common set that were also in the Chat and YouTube sets.

Video	% of Common in Chat	% of Common in YouTube
All videos	68%	73%
Ali G - War	0%	100%
Brothas From the Same Motha	100%	100%
Daughters	50%	50%
Fuggy Fuggy	100%	50%
Gopher Broke	50%	50%
In the Rough	0%	100%
Paddy the Pelican	100%	66%
Pen Pals	100%	50%
Plumber	100%	100%
Tony vs. Paul	100%	50%

Overall, more of the tags in the Common set were found on YouTube than were found in Chat. Since the tags in the Common set were generally highly rated, this disproportion may have unfairly biased the comparison between Chat and YouTube tags. We therefore perform a comparison between three additional tag sources: AllHuman, AllChat, and AllYouTube. These sources were created by averaging the scores for all five tags in their respective category. Thus, in this analysis a single tag may contribute its score to multiple categories. Table 14-4 shows the mean relevancies for each of these new tag sources.

Table 14-4. Relevancies for tag sources, without excluding Common tags. Change scores show the difference in mean (SD) per-tag relevance from their counterparts in Table 14-2.

Tag Source	Mean (SD) per-tag relevance	Change (+/-)	Student's t
AllHuman	2.3 (.56)	+1 (-.09)	A
AllChat	1.8 (.61)	+3 (-.13)	B
AllYouTube	1.6 (.69)	+3 (-.18)	C

The rankings shown in Table 14-4 mirror those of Table 14-2. Tags applied by humans in a laboratory study were of the highest relevancy, followed by tags extracted from chat, and then tags applied by YouTube users. The change scores show that YouTube scores were not unfairly biased from not counting the Common tags; when adding in the Common tags, the mean scores of chat-extracted and YouTube tags both increased by .3.

14.2.5. DISCUSSION

This study addresses the question of whether the set of tags extracted from chat are descriptive of the video. To evaluate the quality of chat-extracted tags, we performed a study to compare the relevancy of chat-extracted tags with tags hand-applied by people in two situations: a controlled laboratory study (to elicit “gold-standard” tags), and on YouTube (to represent real-world taggers).

Overall, tags extracted from chat were significantly more relevant to the videos than tags applied by video uploaders on YouTube. Therefore, we found evidence that relevant tags can be extracted from chat data. The quality of this extraction process was variable, and dependent on the nature of the chat data collected. In three cases, tags from YouTube were rated significantly more relevant than tags extracted from chat. Despite this finding, the highest-rated tags were those that occurred in multiple sets. Therefore, the tag extraction process from chat can augment existing tagging systems to improve tag relevancy.

One explanation for why YouTube tags were rated lower than chat-extracted tags is that tagging on YouTube is performed by the user who uploads a video. Therefore, the quality of the tags is dependent on a single user’s behavior, rather than any “wisdom of the crowd.” It is unclear whether video uploaders are motivated to tag

well or may simply not think of the most relevant tags when uploading a video. In contrast, tags extracted from chat are the collective product of many viewers. Therefore, the quality of these tags may be higher simply because of this fact. However, the startling result is that although viewers chatted about a variety of topics – both related to the video and not – relevant tags were still extracted from their chats. These tags were also collected unobtrusively, as a by-product of the social experience in which they engaged.

A limitation to the generalizability of this study is that the chat data used in this study came from a controlled laboratory environment. It is possible that, in a real-world setting, chats would be more off-topic and less related to the video. In this case, we would expect the tags extracted from chat to be less relevant. Thus, a refinement to the algorithm discussed in this chapter would be to include a text segmentation and classification step between steps (b) and (c) in Figure 14-1 (Shen et al., 2006; Utiyama & Isahara, 2001). The segmentation step creates clusters of chat messages that are topically related to each other, and the classification step determines whether those topics are related to the video or not. Future work is needed to determine whether this refinement can accurately filter out off-topic chat, and boost the relevancy of tags extracted from chat.

14.3. GENERAL DISCUSSION

Tags help people find information online. They are most helpful when their application is efficient, by not having too many tags applied to any particular item, and specific, by having tags be relevant to the items to which they are applied (Chi & Mytkowicz, 2008). Otherwise, when tags are “noisy, ambiguous, and often incorrectly applied, then users will have a hard time finding information in the system” (Chi & Mytkowicz, 2008, p. 82). Therefore, tagging systems should strive to ensure that tags are relevant to the items to which they are applied. Having relevant tags makes it easier for people to find items of interest, either on their own as they browse the tag space, or as part of a recommender system that uses tags to understand relationships between items. In a study of tag-based recommendations, Vig, Sen, and Riedl (2009) found that users better understood why items were recommended to them when they were shown which tags were used to recommend that item, and relevancy of those tags to the item. Thus, it is important to have relevant tags in a tagging system.

The study in this chapter demonstrates that relevant tags can be extracted from noisy chat data. Recall from Chapters 8 and 9 that only approximately 35-40% of participants' chats were focused on the videos themselves; the rest was superfluous to the content of the videos. Using a metric that weights terms based on their uniqueness in a corpus, such as TF-IDF, resulted in tags that were, in many cases, of a higher relevance to the videos than hand-applied tags on YouTube. Thus, the ability to extract relevant tags from raw chat data is beneficial because it can be used to improve the quality of tags and tag-based recommendations.

One subtle point about improving the quality of tags is that tagging systems do not need to exclusively use tags extracted from chat. In the evaluation, the tags that occurred in multiple sources, such as those extracted from chat and applied by hand, were of the highest relevancy. Thus, chat-extracted tags can be used to reinforce existing tagging systems. For example, in the case of a collaborative online video site, video uploaders and viewers can *explicitly* tag videos, and viewers' chats with each other can be used to *implicitly* tag videos. Implicit tagging is desirable because it mitigates problems of under-contribution; even if no one tags a video, relevant tags can still be inferred when viewers chat while watching that video. This strategy of combining explicit and implicit tags is akin to triangulation in the social sciences – using multiple data sources to form a clearer picture of a complex situation.

Another benefit from the ability to extract tags from chat data is that it can help bootstrap the tagging process. Sen et al. (2006) conducted a study in which MovieLens users were asked to apply tags to videos. They found that one reason why some participants did not apply any tags in their study is that they could not think of any tags. Thus, they recommended that tagging systems provide their users with suggestions of tags to apply in the tagging interface. In a collaborative online video site, these tag suggestions can be bootstrapped using the tag extraction method described in this chapter.

Finally, other methods have been used to infer a set of tags for a video using other types of metadata. Siersdorfer, San Pedro, and Sanderson (2009) discuss a method of extracting tags for videos by exploiting redundancy in video content. Their method uses a content-based video analysis to find videos that have overlapping or redundant content. This analysis is used to create a graph of videos, in which edges are present between videos when their content overlaps. This graph can be used to propagate tags between videos, as it exposes videos having similar content.

Siersdorfer et al. conducted a user study of this method, and found that the propagated tags were generally relevant; participants rated the tags between 3 and 4 points on a 5-point scale, with 5 being the highest level of relevancy (Siersdorfer, San Pedro, & Sanderson, 2009, Figure 5).

Future work in this area should be conducted to explore how additional sources of metadata, including both content-based and viewership-based relationships between videos, can be combined with chat data to further increase the relevancy of automatically extracted tags.

14.4. SUMMARY AND CONCLUSIONS

- The process by which tags are extracted from chat involves weighting the tags using a metric (in this chapter, TF-IDF) and selecting the highest-weighted terms.
- The Tag Evaluation study compared tags that were extracted from chat data, tags that were applied by human taggers in a lab study, and tags that were applied by video uploaders on YouTube.
- Tags that occurred in multiple sources were of the highest relevance. In many cases, tags extracted from chat data were significantly more relevant than tags applied on YouTube.
- Tags help people find information online, but only when their application is relevant. The findings in this chapter demonstrate that relevant tags for videos can be extracted from chat data. Thus, chat data can be used to augment the quality of tags in existing tagging systems.

15.

INFERRING VIDEO RATINGS

The occurrence of laughter and evaluative statements in the Cartoon and Text vs. Audio studies (Chapters 8 & 9) suggests there may be a relationship between how one chats and their enjoyment of the video they are watching. For example, the use of positive emotional language such as “happy” or “good” may signal enjoyment of a video, whereas the use of negative emotional language, such as “worthless” or “boring” may signal the opposite. Indeed, the chat coding scheme discussed in Chapter 8 included a category for evaluative language, as people did discuss their feelings toward and liking of the videos they watched.

The presence of evaluative language raises the possibility of automatically inferring one’s enjoyment of a video solely based upon their utterances in chat. The ability to infer a rating for a video from chat data raises an interesting possibility for collaborative online video sites. Descriptive language is generally richer than an N-point rating scale. Therefore, video evaluations that are extracted solely from chat, or generated from a combination of chat and an explicit rating, may be more accurate than those solely collected on an N-point scale. In fact, YouTube – a site that uses a 5-star video rating system – seems to suffer from a bias where unappealing videos are rated highly. This bias can make it difficult for individuals to evaluate videos when most of the videos they see are highly rated.

Improving the accuracy of video ratings can help people find content of interest to them. Ratings also provide feedback to content producers to help them understand whether viewers enjoyed their content; accurate ratings translate to accurate feedback. Finally, accurate ratings are important for recommendation systems because more accurate data translates to higher-relevancy recommendations.

15.1. INFERRING RATINGS FROM LINGUISTIC FEATURES

In this dissertation, I consider the question of whether chat data contains any information that can be used to predict a rating for a video. This task involves using linguistic features to predict a numerical outcome. Thus, linear regression is used to predict a viewer's video rating from the linguistic features present in their chat.

In order to create a set of linguistic features, I employ the Linguistic Inquiry and Word Count (LIWC) dictionary (Pennebaker, Francis, & Booth, 2001). LIWC reduces the feature space of raw linguistic terms to a set of 69 linguistic categories³⁴. Examples of linguistic categories include affect (e.g., "happy," "bitter"), positive emotion (e.g., "happy," "good"), certainty (e.g., "always," "never"), and social processes (e.g., "talk," "friend"). A full listing of LIWC categories is given in Pennebaker, Francis, and Booth (2001).

The categorization of the text chat data from the Cartoon and Text vs. Audio studies into LIWC categories was performed by a Java program I wrote. This program was based on TAWC (Kramer, Fussell, & Setlock, 2004), an open-source Perl implementation of LIWC.

Another set of linguistic features can be constructed using either the raw terms present in the chats or a normalized form of the raw terms (e.g., by applying a stemming or stop word removal algorithm). Such a feature set would be larger than the set defined by LIWC, and this set might provide insight into the specific words or phrases that predict a viewer's rating. However, large feature sets require more data to produce accurate results, and the generalizability of raw terms as features is questionable. For example, knowing that the specific word "good" is a significant predictor of a video's rating is less generalizable than knowing that positive emotional language, which includes both "good" and "awesome," is a significant predictor. Therefore, only LIWC categories were used in this analysis.

15.2. DATA SET

This analysis used the chat data collected from all groups with chat in the Cartoon study, and groups with only text chat in the Text vs. Audio study. There were 27

³⁴ LIWC2001 actually defines 68 categories. I have supplemented the LIWC dictionary with a category for laughter. Details of this addition are discussed in Appendix C.

groups that satisfied these requirement. The resulting data set was comprised of 7,188 chat messages from 88 participants. Table 15-1 shows the general structure of the final data set for the regression analysis.

Table 15-1. Format of the data set for predicting video ratings from linguistic features. The values of each LIWC category were a count of the number of words typed that matched that category. Ratings ranged from 1 (low) to 5 (high). Participants had one row in this data set for each video they watched.

Participant ID	Group	Video ID	{ LIWC Categories }	Rating
----------------	-------	----------	---------------------	--------

15.3. REGRESSION MODELS

As participants watched videos while chatting in groups with other participants, there are several sources of variance that needed to be controlled in the regression. Participant IDs were added to the regression model as a random effect to control for within-participant variance. Second, within-group variance was controlled for by adding the group IDs in which participants chatted to the model. Participant ID was also nested inside of group ID, as participants only belonged to one group. Finally, the video IDs were added as a random effect to control for within-video variance.

Two regression models were created. The affect-only model used only those LIWC categories that related to affective or emotional processes. The full model used all 69 LIWC categories. The intuition behind the affect-only model is that evaluative judgements about a video were often expressed through emotional language. For example, statements such as, “they r so boring” (Cartoon study, participant A1, sic) and, “i like this music” (Cartoon study, participant A2) both expressed evaluations of the video using affective language. The affect-only model was also more parsimonious than the full model, and is thus more desirable.

Table 15-2 reports descriptive statistics and regression results for both models. Videos had a mean (SD) rating of 3.1 (1.4) points on a 5-point scale. In the table, the “M (SD)” column gives the mean (SD) number of words spoken per person for the given category. Standard errors (SE) on regression coefficients are listed in parenthesis. The table reports both marginally significant effects ($p < .10$) and significant effects ($p < .05$) to help us understand the general trends in the data set.

Table 15-2. Descriptive statistics of per-participant linguistic features and regression models for predicting video ratings from these features. Mean values are of the number of words spoken in each category. Unstandardized regression coefficients are reported for each model. Standard errors (SE) on regression coefficients are given in parenthesis. Level of significance for the coefficients are reported as (*) $p < .10$ and () $p < .05$. The table data pertains to 88 participants with only text chat in the Cartoon and Text vs. Audio studies.**

	M (SD)	Affect-Only	Full Model
Intercept		3.06** (.18)	3.20** (.19)
Control			
Video ID	7 vids. (C); 8 vids. (TA)		random effect
Group[Participant ID]	18 grps. (C); 12 grps. (TA)		random effect
Participant ID (random effect)	57 Ps (C); 36 Ps (TA)		random effect
Standard Linguistic Dimensions			
Total pronouns (Pronoun)	3.86 (4.01)		.01 (.05)
1st person singular (I)	1.68 (1.93)		zeroed
1st person plural (We)	.16 (.49)		-.28 (.30)
Total first person (Self)	1.84 (2.12)		.05 (.07)
Total second person (You)	.50 (.98)		-.02 (.28)
Total third person (Other)	.68 (1.24)		-.07 (.28)
Negations (Negate)	.49 (.84)		-.12 (.07)*
Assents (Assent)	.53 (.84)		-.09 (.07)
Articles (Article)	1.61 (1.97)		-.0008 (.04)
Prepositions (Preps)	2.12 (2.47)		.07 (.04)*
Numbers (Number)	.37 (.71)		.06 (.08)
Psychological Processes			
Affective or Emotional Processes (Affect)	2.22 (2.08)	.59 (.42)	.67 (.43)
Positive Emotions (Posemo)	1.54 (1.56)	-.48 (.42)	-.52 (.43)
Positive feelings (Posfeel)	.54 (.85)	-.04 (.08)	-.08 (.09)
Optimism and energy (Optim)	.14 (.38)	-.18 (.14)	-.16 (.16)
Negative Emotions (Negemo)	.69 (1.04)	-.89** (.41)	-.91 (.43)**
Anxiety or fear (Anx)	.07 (.27)	.30 (.19)	.37 (.20)*
Anger (Anger)	.27 (.59)	.05 (.12)	.01 (.14)
Sadness or depression (Sad)	.08 (.31)	.21 (.16)	.002 (.19)
Cognitive processes (Cogmech)	1.79 (2.17)		-.10 (.07)
Causation (Cause)	.24 (.57)		.17 (.12)
Insight (Insight)	.51 (.85)		.01 (.10)
Discrepancy (Discrep)	.48 (.86)		.13 (.10)
Inhibition (Inhib)	.10 (.35)		.17 (.19)
Tentative (Tentat)	.76 (1.13)		-.15 (.06)**
Certainty (Certain)	.26 (.57)		.03 (.10)
Sensory and Perceptual Processes (Senses)	.84 (1.11)		.05 (.28)

	M (SD)	Affect-Only	Full Model
Seeing (See)	.41 (.71)		-.02 (.29)
Hearing (Hear)	.30 (.63)		.22 (.30)
Feeling (Feel)	.07 (.28)		-.30 (.33)
Social Processes (Social)	2.34 (2.89)		.11 (.20)
Communication (Comm)	.42 (.77)		-.26 (.23)
Other references to people (Othref)	1.37 (1.99)		-.04 (.34)
Friends (Friends)	.02 (.16)		-.77 (.39)**
Family (Family)	.13 (.45)		-.02 (.25)
Humans (Humans)	.31 (.65)		.01 (.23)
Relativity			
Time (Time)	.73 (1.16)		-.04 (.06)
Past tense verb (Past)	.88 (1.27)		-.13 (.06)**
Present tense verb (Present)	3.33 (3.20)		-.04 (.04)
Future tense verb (Future)	.24 (.59)		.01 (.10)
Space (Space)	.54 (.93)		-.0003 (.08)
Up (Up)	.28 (.63)		-.16 (.10)
Down (Down)	.06 (.28)		.20 (.20)
Inclusive (Incl)	1.37 (1.61)		-.02 (.04)
Exclusive (Excl)	.91 (1.35)		.06 (.06)
Motion (Motion)	.21 (.51)		-.07 (.11)
Personal Concerns			
Occupation (Occup)	.39 (.73)		-.13 (.28)
School (School)	.09 (.37)		.15 (.32)
Job or work (Job)	.10 (.35)		-.14 (.28)
Achievement (Achieve)	.21 (.48)		.12 (.29)
Leisure activity (Leisure)	.54 (.93)		1.04 (.56)*
Home (Home)	.09 (.33)		-.94 (.55)*
Sports (Sports)	.06 (.28)		-.91 (.60)
Television and movies (TV)	.22 (.56)		-.93 (.54)*
Music (Music)	.17 (.55)		-1.15 (.55)**
Money and financial issues (Money)	.13 (.42)		.09 (.13)
Metaphysical issues (Metaph)	.09 (.32)		-.18 (.29)
Religion (Relig)	.06 (.25)		.09 (.37)
Death and dying (Death)	.03 (.18)		zeroed
Physical states and functions (Physcal)	.42 (.84)		.29 (.27)
Body states, symptoms (Body)	.22 (.61)		-.15 (.25)
Sex and sexuality (Sexual)	.13 (.41)		.08 (.26)
Eating, drinking, dieting (Eating)	.08 (.32)		-.33 (.26)
Sleeping, dreaming (Sleep)	.02 (.14)		-.62 (.44)
Grooming (Groom)	.03 (.18)		-.61 (.35)*
Experimental Dimensions			
Swear words (Swear)	.09 (.36)	.21 (.16)	.17 (.18)
Nonfluencies (Nonfl)	.07 (.27)		.05 (.19)
Fillers (Fillers)	.001 (.04)		-.51 (1.28)
Laughter (Laugh)	.82 (1.42)	.06* (.04)	.07 (.04)*

	M (SD)	Affect-Only	Full Model
R ² (R ² adjusted)		.37 (.36)	.44 (.38)
Root mean squared error (RMSE)		1.17	1.16

Overall, the most popular words typed in chat were in linguistic categories such as pronouns (e.g., “our,” “they”; M [SD] = 3.84 [4.00] occurrences per participant), present tense verbs (e.g., “is,” “be”; M [SD] = 3.21 [3.10] occurrences per participant), prepositions (e.g., “on,” “from”; M [SD] = 2.11 [2.46] occurrences per participant) and self-references (e.g., “me,” “I”; M [SD] = 1.84 [2.12] occurrences per participant). Popular affective categories were affect (e.g., “happy,” “afraid”; M [SD] = 1.85 [1.81] occurrences per participant) and positive emotions (e.g., “agree,” “best”; M [SD] = 1.27 [1.34] occurrences per participant). Cognitive process words were also popular (e.g., “cause,” “know”; M [SD] = 1.71 [2.08] occurrences per participant).

15.3.1. AFFECT-ONLY MODEL

The affect-only model used the following LIWC categories as predictive features: affective or emotional processes (Affect), positive emotions (Posemo), positive feelings (Posfeel), optimism and energy (Optim), negative emotions (Negemo), anxiety or fear (Anx), anger (Anger), sadness or depression (Sad), curse words (Swear), and laughter (Laugh).

The affect-only model had an adjusted $R^2 = .36$ and a $RMSE = 1.17$. Negative emotional language was a significant predictor of video ratings ($b = -.89$, $p < .05$). As expected, the coefficient was negative; thus, each additional negative emotional word spoken in chat decreased the predicted video rating by .89 points.

Laughter was another significant predictor of video ratings, and it had a positive coefficient ($b = .06$, $p < .10$). Thus, each additional “haha” or “hehe” in chat increased the predicted video rating by .06 points.

15.3.2. FULL MODEL

The full model used all but two LIWC categories as predictive features. Two categories are excluded from the full model because they were found to be linear

combinations of other categories. They were I (Self - We) and Death (Metaph - Relig). The final model had an adjusted $R^2 = .38$ and a RMSE = 1.16.

As in the affect-only model, laughter was a significant predictor of video ratings, with a positive coefficient ($b = .07$, $p < .10$). Several other linguistic categories were also significant predictors. In the positive direction, more prepositions (e.g., “on,” “to”) and more language about leisure activities (e.g., “dance,” “drums”) were associated with higher video ratings. Counterintuitively, more anxious language (e.g. “nervous,” “tense”) was also associated with higher video ratings.

In the negative direction, negations (e.g., “not,” “never”), language about negative emotions (e.g., “hate”), tentative language (e.g., “maybe,” “perhaps”), and past tense verbs were associated with lower video ratings. Further, language that may be classified as “off-topic,” such as talking about one’s friends (e.g., “boyfriend,” “buddy”), the home (e.g., “house,” “kitchen”), and TV or music (e.g., “sitcom” or “song”) were negatively associated with video ratings.

15.4. DISCUSSION

The two regression models explained about 36-38% of the variance in video ratings³⁵. The small difference in the adjusted R^2 values between the two models was consistent with the prediction that video evaluations are often expressed with affective language: adding predictors for non-affective language (and controlling for the number of features added) only increased explained variance by about 2%.

As for the quality of the regression models, the standard deviations of the residual error (the error between predicted and actual values) were 1.17 and 1.16 for the affect-only and full models, respectively. Put into perspective, this means that a prediction made by these models will be “off” by a little more than a point. Given that the only data used to make these predictions were the linguistic markers present in each participant’s chat, these models are generally predictive of video ratings. As they explain only 36-38% of the variance in video ratings, they can also be improved.

³⁵ These R^2 values were adjusted for the number of prediction terms in the regression. This adjustment added a penalty to avoid over-fitting from having a large number of prediction terms (affect-only: 13 prediction terms, full model: 70 prediction terms).

In both models, negative emotional language was a significant predictor of video ratings. The effect of this language was fairly strong even though its occurrence was relatively infrequent, compared to other types of language (on average, participants spoke .69 Negemo words per video). The utterance of one negative word (e.g., saying “this is *bad*”) decreased the predicted video rating by about .89 - .91 points on the 5-point rating scale. Interestingly, the effect of positive emotional language (e.g., saying “this is *good*”) did not have a contrary effect, even though positive emotional language was more frequently used (on average, participants spoke 1.54 Posemo words per video). One explanation of this finding is that participants may have expressed their positive feelings toward the videos using alternative forms, such as laughter or smilies. Indeed, laughter was a significant predictor in both models. Single utterances of laughter increased predicted video ratings by around .06 - .07 points. Although these beta coefficients seem small, they predict a .6 - .7 point increase in a video’s rating for an individual who laughs 10 times during a video.

The results of this analysis support the hypothesis that a viewer’s usage of linguistic terms is reflective of their enjoyment of the video they are watching. In fact, about 38% of the variance in people’s ratings of videos was explained solely by what they said while watching those videos. To improve the quality of the inferred video ratings, future research should be conducted to examine alternative linguistic features beyond those provided by the LIWC dictionary. These features can be combined with other information about viewers, such as their ratings of different genres, directors, or actors, in order to compute more accurate video ratings.

In addition, the LIWC feature set may be able to provide a richer interpretation of a viewer’s experience as they watch a video beyond just a numerical rating. LIWC contains multiple categories for affect, such as positive emotions, positive feelings, optimism, anger, and sadness. These qualities can be used to evaluate a richer spectrum of enjoyment than a simple numerical video rating. Future research is needed to determine whether the expression of these categories are accurately representative of viewers’ emotional responses to a video.

15.5. SUMMARY AND CONCLUSIONS

- This chapter demonstrates a relationship between a viewer's use of language in chat and his or her enjoyment of a video. This relationship can be used to improve the quality of ratings in collaborative online video sites.
- The Linguistic Inquiry and Word Count (LIWC) dictionary is used to classify terms uttered in chat into linguistic categories. The LIWC dictionary was extended to support textual laughter such as "haha" and "lol."
- Linear regression models used LIWC categories as features to predict video ratings. These models accounted for 36-38% of the variance in video ratings. Video predictions were off by about 1 point on the 5-point rating scale, demonstrating that the models were both predictive and improvable.
- Negative emotional language was associated with lower video ratings, and laughter was associated with higher video ratings. Off-topic chat, such as talking about one's friends, home, TV, or music, were negatively associated with video ratings.
- These results of this analysis highlight features of chat that can be used to create more accurate metrics of viewers' enjoyment of videos. Future work should be conducted to consider richer interpretations of enjoyment beyond numerical ratings.

16.

LEARNING ENJOYMENT PROFILES

As seen in the Cartoon and Text vs. Audio studies (Chapters 8 & 9), laughter comprises about 10-17% of viewers' chat while they watched videos together. From reading chat transcripts, this laughter often seemed to coincide with points in the video that participants found enjoyable or funny. This observation leads to a natural question: if people laugh during the parts they enjoy or find funny, can aggregating laughter across multiple viewers expose a general profile of enjoyment for a video?

Knowing which parts of a video are enjoyable can be useful for online video sites. For example, as longer video content becomes more readily available, labeling the parts other viewers enjoyed could help users find interesting content or funny clips. These labeled sections also provide finer-grained feedback to content creators on their works.

This chapter considers two research questions regarding enjoyment profiles. First, can they be built by aggregating laughter across viewers, or is laughter completely uncorrelated across different viewers? Second, are the profiles extracted from chat data reflective of profiles created by human raters?

16.1. PRIOR WORK

Shamma et al. (2007) posited that community activity, such as patterns of chatting, watching, pausing, and rewinding, could be used to understand viewers' engagement and interest in video content. In a case study of aggregated activity for a video, they noticed that seek behaviors were more frequent at the beginning of the video and chat behaviors were more frequent during the later parts of the video.

After seeing these patterns, they concluded that the later parts of the video were more interesting or enjoyable to viewers than the beginning parts. However, these observations were the only extent to which Shamma et al. explored the idea of learning about videos through patterns of chat activity.

Following up on this idea, the CollaboraTV system (Nathan et al., 2008) allowed viewers to explicitly vote thumbs up / thumbs down while watching television shows. These votes were then used to create moment-by-moment profiles of interest over time, similar to those examined in this chapter. One limitation of this method was that these votes must be explicitly cast by viewers. In an evaluation study of CollaboraTV, Nathan et al. (2008) only 24 instances of marking interest points were made by 16 viewers. The technique for generating enjoyment profiles discussed in this chapter uses an unobtrusive measure of enjoyment – the presence of laughter – that avoids the requirement that viewers explicitly register their enjoyment.

San Pedro, Kalnikaite, and Whittaker (2009) used video content redundancy in an online video system to determine the important sections in each video clip. Their method of determining the level of redundancy present across video clips was similar to that used by Siersdorfer, San Pedro, and Sanderson (2009). The general idea was that the importance of the scenes in a video could be determined by measuring how often those scenes were uploaded in other video clips. However, this method only judges importance, and not necessarily enjoyment. For example, many people have uploaded clips of the 9/11 attacks to YouTube; this redundancy merely signals that those clips are important, and not that they are enjoyable to watch.

Miyamori, Nakamura, and Tanaka (2005) have previously examined the problem of identifying “high points” in a video – the points where enjoyment is highest – by analyzing live text chats. They used two features in chat data to measure enjoyment. First, they used a small dictionary of 29 words that represented enjoyment and disappointment. This dictionary included terms like “amazing,” “wow,” (enjoyment) and “sigh” (disappointment). Second, they used a set of regular expressions that analyzed ASCII art and classified it as either expressing viewers’ enjoyment or disappointment with the video. Using these features, Miyamori et al. showed that viewers’ enjoyment could be computed over time. They also showed that chat messages could be accurately classified as expressing “enjoyment” (F-statistic = .942); classifying chat messages as “disappointment” was slightly less accurate (F-statistic = .825).

One limitation of the method used by Miyamori et al. is that it exhibits a cultural bias. In the Cartoon and Text vs. Audio studies, the same types of ASCII art described by Miyamori et al. were not observed. Thus, accuracy of their enjoyment metric would likely decrease for a non-Japanese viewership. Further, Miyamori et al. did not consider shared laughter, which comprised a significant portion of the chats in the Cartoon and Text vs. Audio studies. The lack of consideration of laughter may have reflected cultural differences between how American and Japanese viewers express themselves in chat. Therefore, the task of learning enjoyment profiles from chat is re-evaluated in this dissertation.

16.2. AGGREGATING LAUGHTER

Laughter is an important aspect of the collaborative online video watching experience. As we saw in the Cartoon and Text vs. Audio studies, laughter comprised between 10-17% of viewers' chat while they watch videos. Further, the analysis in Chapter 15 showed that laughter was a significant and positive predictor of one's enjoyment of a video: more laughter corresponded to a higher video rating. Therefore, laughter serves as a signal for whether a viewer has enjoyed a video.

Laughter carries more information in it than just its presence. Assuming that a viewer laughs at the specific instances in the video that he or she finds enjoyable (as opposed to laughing at random times), the occurrence of laughter can also be used to denote the particular moments in the video the viewer found enjoyable or funny. On the surface, this assumption seems to be reasonable. We tend not to laugh randomly while watching television or a movie, unless perhaps we are thinking of something else at the time. However, this assumption may not be reasonable in the collaborative online video case. When watching with others, a viewer's laughter may either be in response to something funny in the video or to something funny said in the chat. This extraneous laughter can be thought of as noise in the system: if the goal is to find the portions of a video a viewer most enjoyed, laughter that occurred from something funny in the chat will not help.

One solution to this problem is to aggregate a viewer's laughter across repeated views of a video. This method results in an accurate enjoyment profile for an individual viewer, as long as he or she consistently laughs at the points in the video he or she most enjoys, and all other laughter (e.g., from discussing non-video related

topics) is randomly distributed. However, this method is unrealistic as it requires viewers to repeatedly watch the same video in order to collect enough data to create an accurate profile.

An alternative method for building an enjoyment profile is to aggregate laughter across different viewers of a video. In this case, laughter elicited from the video will be correlated among viewers, and laughter produced in response to off-topic chat will be randomly distributed. Thus, the enjoyment profile should exhibit a clear peaks and valley distribution: peaks at the locations of the most enjoyable parts, and valleys representing noise. Figure 16-1 shows an example of a hypothetical enjoyment profile with these features.

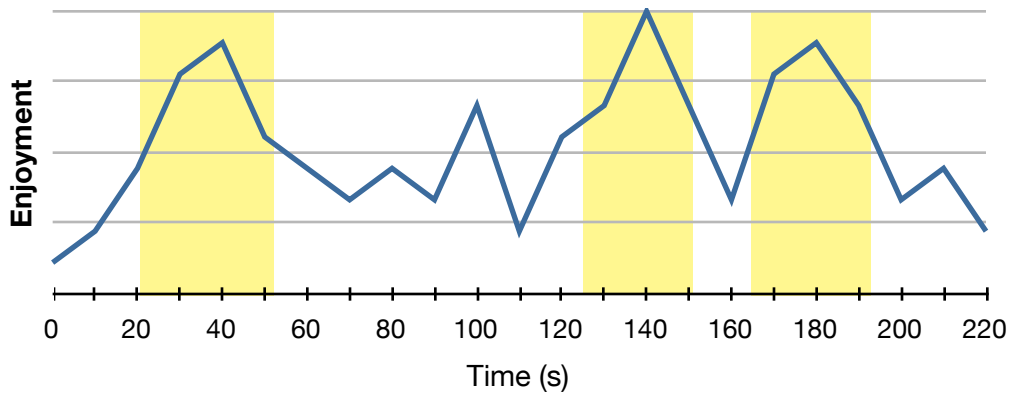


Figure 16-1. Hypothetical enjoyment profile. Shaded regions indicate the locations of the high points.

The goal of this chapter is to understand whether enjoyment profiles created from aggregated laughter exhibit the characteristics of Figure 16-1: are the peaks clearly visible? How much data is needed to make a clear picture? How fine-grained can the profile's resolution be? Are enjoyment profiles extracted from chat representative of enjoyment profiles created by human raters?

16.3. EXTRACTING ENJOYMENT PROFILES FROM CHAT DATA

Extracting an enjoyment profile for a video from chat data involves creating a histogram of laughter over time by counting how much laughter occurred in each R-second period during the video, where R is the resolution of the histogram. After this histogram is created, a set of high points can be found by looking for the entries in the histogram with the highest value. These points are the locations at which aggregated laughter was highest.

This procedure makes two simplifying assumptions. The first is that the enjoyable parts of the video are strictly aligned to R-unit intervals, beginning at zero. For example, with a resolution of 10 seconds, we might believe that the first high point in Figure 16-1 is located between 20 and 50 seconds. However, a resolution of 5 seconds might reveal a bimodal distribution: one peak from 20-30 seconds, and another between 45-50 seconds. Therefore, the resolution at which laughter is aggregated is important to the interpretation of the enjoyment profile. Too large of a resolution will hide fine details in the profile, but too small of a resolution may cloud the interpretation of the profile or make the task of locating high points more difficult (e.g., a profile with a resolution of 1 second might look completely flat). A comparison between a 15-second and 5-second resolution is shown in Figures 16-2 (e) and (f). The 15-second resolution shows a trend of high enjoyment during the early and middle parts of the video (30s-165s), with a drop in enjoyment toward the end (after 165s). The 5-second resolution shows much more variance over time in enjoyment, with two peaks at 65s and 170s. In this case, the general trend is clearer with a lower resolution, and the specific high points are clearer with a higher resolution. This observation suggests that algorithms that automatically identify and label the best parts of video should consider multiple resolutions when aggregating momentary enjoyment data.

The second assumption is that the time at which laughter was uttered corresponds to the exact moment in the video that prompted the laughter. This assumption is not likely valid, as there is usually a delay between stimulus and response when laughing. This assumption can be relaxed by adjusting the time points in the laughter histogram. One way this adjustment can be made is by modeling the amount of time between stimulus (funny part) and response (laughter) and correcting for it in the laughter's timestamp. In fact, Miyamori, Nakamura, and Tanaka (2005) performed such a correction by modeling the amount of time it took for a user to type in a chat message. A simpler approach is to simply subtract a constant (e.g., 2 or 3 seconds) from each laughter timestamp as an approximation to that model. However, both of these corrections simply shift the distribution of timestamps to the left without significantly altering the shape of the distribution, as the correction is generally smaller (e.g., 2 to 3 seconds) than the resolution of the graph (e.g., 5 to 10 seconds). Therefore, no timestamp corrections are performed in the analysis presented in this chapter.

16.3.1. CHAT-EXTRACTED ENJOYMENT PROFILES OF THE STUDY VIDEOS

Enjoyment profiles were extracted from the chat data for each of the 15 videos used in the Cartoon and Text vs. Audio studies. All chat data was used in this analysis. Thus, the videos in the Text vs. Audio study represent situations in which viewers used both textual and auditory chat. The videos in the Cartoon study represent situations in which viewers only used text chat. This distinction allows us to contrast between situations in which a larger or smaller amount of data is available, as there was an order of magnitude more auditory laughter than textual laughter in the data set. Table 16-1 details the mean number of instances of laughter per video for the videos in each study.

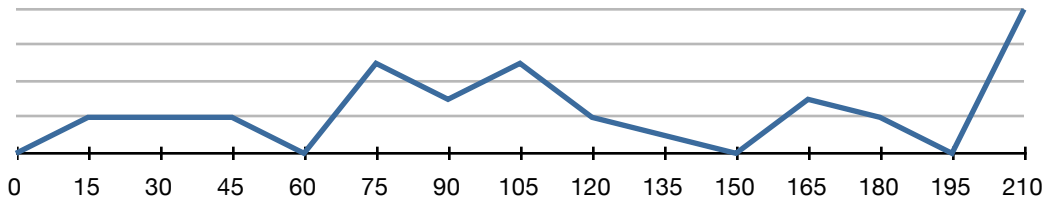
Table 16-1. Amount of laughter for the videos across all groups in the high and low data conditions. Standard deviations are listed in parentheses.

Condition	# videos	Mean (SD) # laughter instances per video	Mean (SD) # groups with any laughter
Low data (Cartoon)	7	45.8 (13.3)	11.7 (2.3) of 20
High data (Text vs. Audio)	8	313.7 (96.9)	31.0 (2.6) of 36

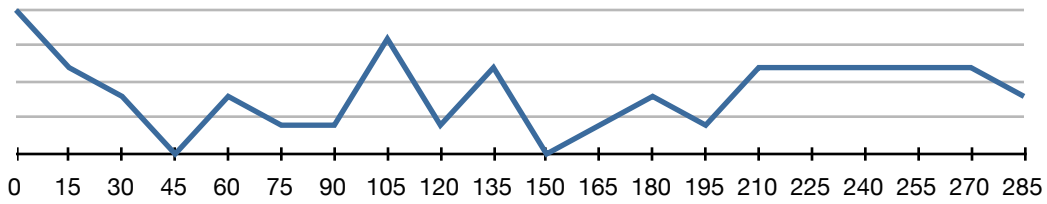
In the low-data condition, there was an average of 46 instances of laughter per video. Not all groups laughed during each video, and in this condition, an average of 12 groups (60%) had any laughter for each video. In the high-data condition, which included out-loud laughter, there was an order of magnitude more laughter per video ($M = 314$ instances). In addition, a greater proportion of groups laughed during these videos ($M = 86\%$).

Figure 16-2 shows the chat-extracted enjoyment profiles for six videos. The set of high points in the video, taken as the points in the video where laughter was highest, are listed as well. These particular profiles were selected to highlight different trends in the profiles. Figures 16-2 (a) and (b) show profiles for two videos in the Cartoon study, built using only laughter from text chat. These figures use a resolution of 10 seconds. Figures 16-2 (c) and (d) show profiles for two videos in the Text vs. Audio study, built using laughter from both text and voice chat. Figures 16-2 (e) and (f) compare between 5-second and 15-second resolutions for the same video. This comparison highlights how aggregation can cause a shift in the interpretation of an enjoyment profile.

Cartoon study videos (text chat only; low-data condition)

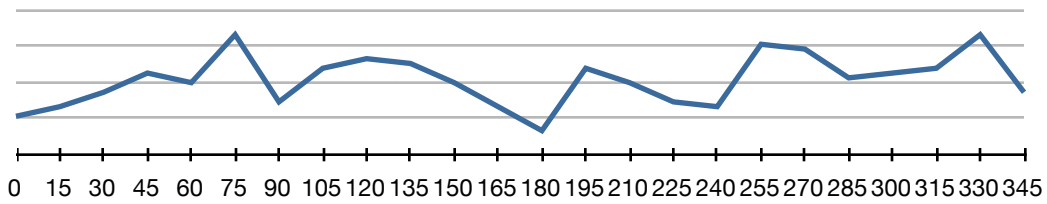


(a) *Emerge* (35 laughter instances). High points: 75s-105s, 210s.

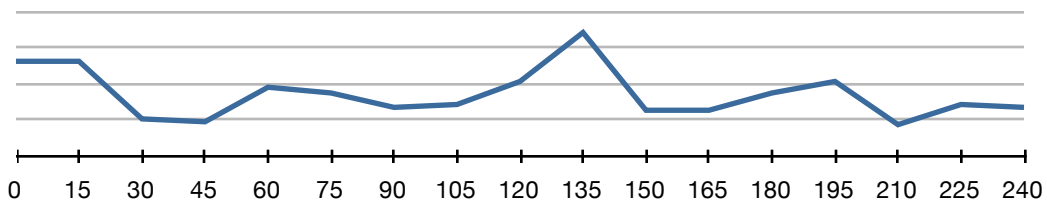


(b) *In the Rough* (43 laughter instances). High points: 0s, 105s, 135s, 210s-270s.

Text vs. Audio study videos (text and voice chat; high-data condition)

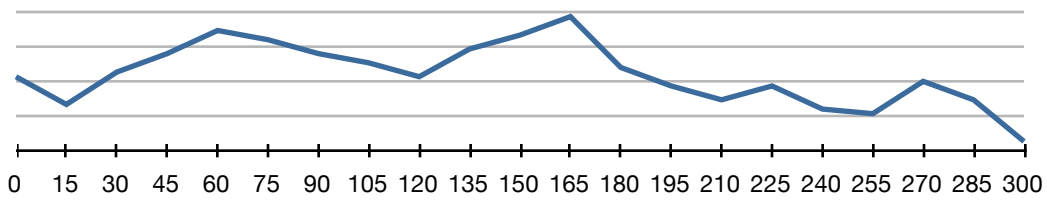


(c) *Ali G - War* (374 laughter instances). High points: 75s, 120s-135s, 195s, 255s-270s.



(d) *Gopher Broke* (367 laughter instances). High points: 0s-15s, 120s-150s.

Comparison between 15-second and 5-second resolutions (text and voice chat; high-data condition)



(e) *Tony vs. Paul* (348 laughter instances). 15-second resolution. High points: 60s, 165s..

or the latter. Therefore, the chat-extracted profiles suffer from some amount of noise due to laughter unrelated to the videos. Noise may cause one to misinterpret the profile or incorrectly label the locations of the most enjoyable parts.

The Best Part Labeling study was conducted to evaluate the accuracy of the chat-extracted enjoyment profiles. In this study, human raters watched the videos and manually labelled their favorite parts in the video. These hand-created profiles were then compared with the profiles extracted from chat. Although it is tempting to assume that the hand-created profiles constitute a “gold standard” of video enjoyment, we must be careful of the fact that individual differences cause people to find different parts of a video enjoyable. Therefore, the degree to which the hand-created profiles exhibit consensus is also examined.

16.4.1. PARTICIPANTS

Twenty participants were recruited through the CBDR web site and word of mouth. Participants were generally college-aged (18-24) and spoke English as their native language. Participants were paid \$15 for their time. Participation in this study took approximately one and a half hours.

16.4.2. METHOD

Participants watched the 15 videos listed in Chapter 5. Immediately after watching each video, they filled out a survey that asked them to rate the video (1 - 5 stars), identify up to four portions of the video they enjoyed the most, and tag the video³⁶. Participants watched the videos in a random order to guard against order effects.

To identify their favorite parts, participants were asked to provide the time indices of the beginnings and ends of the parts. Participants were allowed to take notes on while watching the video to record the locations of their favorite parts. They were also allowed to scrub through each video after watching it to re-locate their favorite parts. To make their task easier, participants were told to report their favorite parts to the nearest five seconds. Thus, instead of having to reason whether their favorite part began at 12 or 13 seconds, they could simply round down to 10 seconds.

³⁶ Tags were collected in this study for use in the Tag Evaluation study (Chapter 14). Tags were explained to participants as “short words or phrases that describe the video”. Participants were allowed to apply as many tags as they wanted to each video.

Participants were told that their favorite parts could be of any length. In addition, they were instructed to not report any favorite parts for a video if they did not enjoy any portion of that video. Finally, for each favorite part reported, participants were asked why they selected that part as their favorite.

16.4.3. RESULTS

To create enjoyment profiles from participants' reported time ranges, an aggregation method similar to the one for creating chat-extracted profiles was used. For each R-second histogram bucket, we counted the number of people who had a time range that fell into that bucket. For example, a time range of "0:50-1:05" (50 seconds to 1 minute and 5 seconds) contributed a "vote" to the histogram buckets representing 50-55 seconds, 55-60 seconds, and 60-65 seconds (using 5-second buckets). Therefore, each bucket in the hand-created enjoyment profile represented the number of people who felt that that bucket was enjoyable.

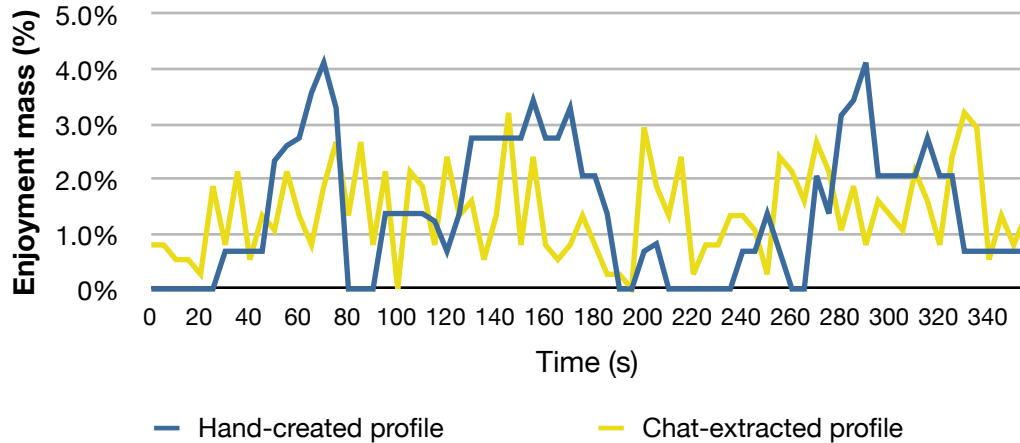
Overall, the 20 participants reported an average of 1.4 (SD = .83) favorite time ranges per video. Participants reported more favorite time ranges when they rated the video higher ($r = .48$, $p < .0001$). Several reasons for why a time range was marked as a favorite included visual or auditory characteristics of the video, comical or humorous qualities, plot and plot twists, and characters. Examples of each of these are presented in Table 16-2. Note that this table is merely representative of some of the reasons given; it is not meant to be comprehensive.

Table 16-2. Several categories of reasons for why participants labeled specific parts as being their favorite parts. Commenters are labelled with their study ID, and comments are reproduced in their original form.

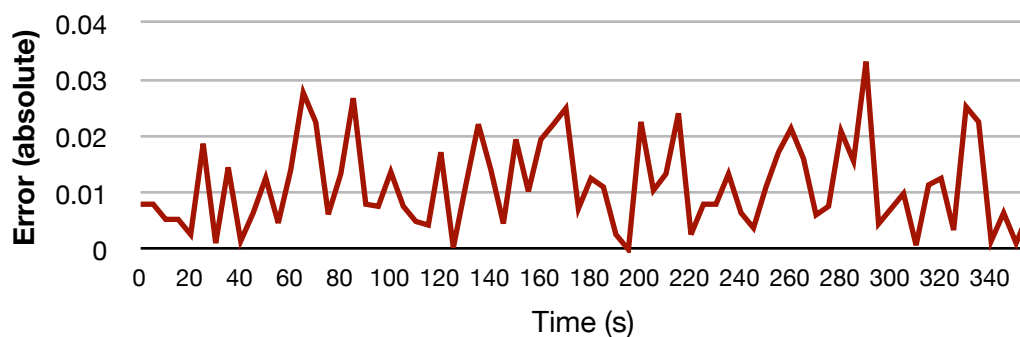
Category	ID	Video	Comment
Visual or auditory characteristics of the video	BP8	Emerge	"music is great"
	BP12	Powaqqatsi	"photography great contrast"
Comical or humorous qualities	BP7	Tony vs. Paul	"It's funny enough that I can't help but laughing."
	BP20	Daughters	"The entire video was HILARIOUS"
Plot and plot twists	BP5	Gopher Broke	"unexpected ending."
	BP4	Plumber	"he is actually a plumber!!"
Characters	BP1	Fuggy Fuggy	"ninja pig"
	BP5	Ali G - War	"sometimes it funny when people are being purposely ignorant"

The two research questions pertaining to hand-created enjoyment profiles were whether they correlated with chat-extracted profiles, and how much inter-rater consensus was present.

Agreement between hand-created and chat-extracted profiles. Figure 16-3a shows normalized plots of both types of enjoyment profiles for “Ali G - War,” using a resolution of 5 seconds. Each y-value on the plot represents the percentage of enjoyment mass for the corresponding time bucket. For example, a value of .0375 at $t = 70s$ corresponds to 3.7% of the enjoyment mass. A visual inspection shows that there is not much agreement between the two enjoyment profiles. Figure 16-3b shows the magnitude of error at each time point. Again, visual inspection shows that there is significant disagreement in enjoyment ratings for many of the time points.



(a) Profile comparison. The blue line (darker) is the hand-created profile. The yellow line (lighter) is the chat-extracted profile.



(b) Absolute error in enjoyment mass between the hand-created and chat-extracted enjoyment profiles. Lower values indicate higher agreement.

Figure 16-3. (a) Comparison of the hand-created profile with the chat-extracted profile for “Ali G - War”. (b) Absolute error between the profiles.

$$\begin{aligned}
 CP_t &= \text{mass of chat extracted profile at time } t \\
 HP_t &= \text{mass of hand created profile at time } t \\
 E_t &= |CP_t - HP_t| \\
 E &= \sum_t E_t \\
 A &= \frac{2 - E}{2} \cdot 100
 \end{aligned}$$

Figure 16-4. Equations for computing percentage agreement (*A*) between chat-extracted and hand-created profiles.

To quantify the differences between the hand-created and chat-extracted profiles, we compute a percentage agreement metric. Equations for this computation are shown in Figure 16-4. Let CP_t and HP_t be the mass of the enjoyment distribution at time t for the chat-extracted and human-created profiles, respectively. Let E_t be the absolute error between these profiles at time t (shown in Figure 16-3b). The agreement error between the profiles (E) is the sum of the individual errors over time. Agreement error E has a range of 0 (perfect agreement) to 2 (perfect disagreement). Thus, the percentage agreement (A) can be computed as shown in Figure 16-4. The percentage agreement for each video is given in Table 16-3, along with the correlations of enjoyment mass over time between the profiles.

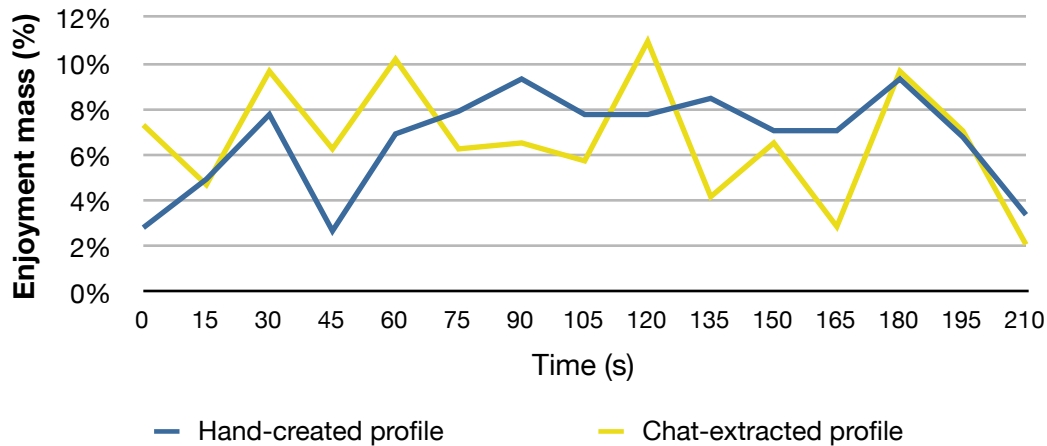
Table 16-3. Percentage agreement and correlations of enjoyment mass over time for both 5-second and 15-second resolutions. (*) $p < .05$ for the correlations.

Video	Agreement (%)		Correlations	
	5 sec	15 sec	5 sec	15 sec
Cartoon – low data condition				
Emerge	42.4	46.6	-.01	-.14
Fuggy Fuggy	55.8	69.6	.29*	.38
In the Rough	43.1	59.0	-.02	-.05
Pen Pals	45.9	55.4	.09	.09
Penguin’s Christmas	36.4	46.8	.27*	.43
Plumber	44.4	55.2	-.0043	.09
War Photographer	37.0	58.7	.07	.11
Text vs. Audio – high data condition				
Ali G - War	58.9	67.1	.11	.16
Brothas From the Same Motha	60.9	65.0	-.07	-.11
Daughters	75.9	83.0	.05	.33
Gopher Broke	56.8	65.9	.04	.25

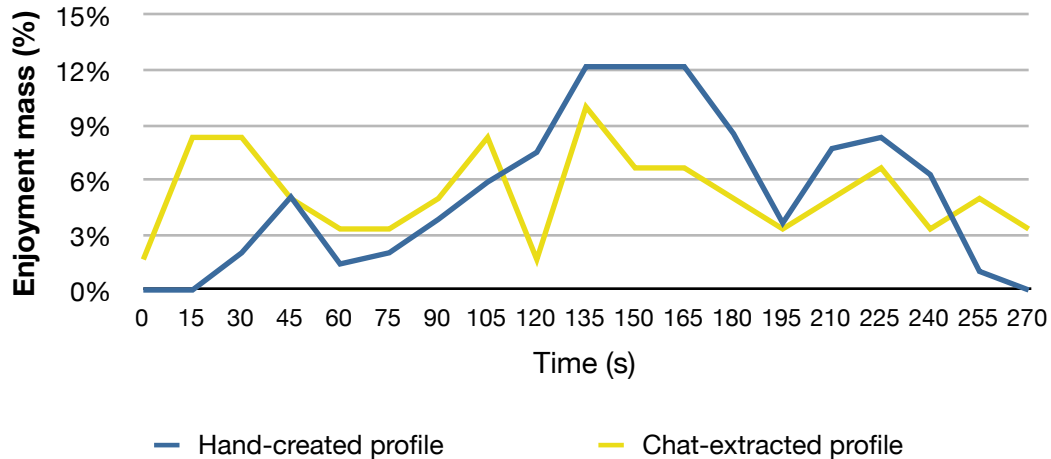
Video	Agreement (%)		Correlations	
Paddy the Pelican	39.4	53.0	-.06	-.11
Powaqqatsi	67.1	76.5	.05	.15
Tea	46.6	58.5	.14	.25
Tony vs. Paul	65.2	78.1	.08	.25

Agreement between the hand-created and chat-extracted profiles ranged from about 36% to 76% using a 5-second resolution, and from about 46% to 83% using a 15-second resolution. Agreement tended to be higher for those videos with more laughter (high-data condition). Overall, profiles tended to not to correlate very well at either resolution, with notable exceptions being “Fuggy Fuggy” ($r = .29, p < .05$) and “Penguin’s Christmas” ($r = .27, p < .05$).

To understand agreement visually, Figure 16-5 shows the hand-created and chat-extracted profiles for “Daughters,” the video with the highest profile agreement as measured by percent agreement, and “Fuggy Fuggy,” the video with the highest profile agreement as measured by correlation. Both graphs are shown with a 15-second resolution.



(a) Enjoyment profiles for “Daughters” at the 15-second level. Profile agreement is 83.0% and correlation is $r = .33$.



(b) Enjoyment profiles for “Fuggy Fuggy” at the 15-second level. Profile agreement is 69.6% and correlation is $r = .38$.

Figure 16-5. Enjoyment profiles for (a) “Daughters” and (b) “Fuggy Fuggy.” The blue line (darker) is the hand-created profile; the yellow line (lighter) is the chat-extracted profile.

In both cases, similar trends can be seen between the hand-created profiles and the chat-extracted profiles. For example, in “Daughters,” both profiles captured a peak-dip-peak pattern in enjoyment in the 30s-60s range. In “Fuggy Fuggy,” a similar pattern is seen for the 30s-105s range, although the chat-extracted profile picked up the peak trend earlier than the hand-created profile.

Consensus among raters. The second research question in this study is whether individual raters expressed much consensus for where they felt the best parts of the video were located. If raters do not generally not agree on which parts they found most enjoyable, then the aggregation method used to construct enjoyment profiles from chat data must be reconsidered. For example, a clustering step may be required to group viewers based on their video preferences (i.e., collaborative filtering). After clustering, aggregated profiles could be built each cluster. In this way, chat-extracted profiles could account for individual differences in enjoyment.

To measure the level of consensus, we need to know whether participants tended to mark the same sections of the videos as their favorite parts. Consensus in favorite-part selection manifests itself as higher peaks and lower valleys in the enjoyment profile distribution. However, standard measures of the spread and peakedness of a distribution, such as standard deviation and kurtosis, cannot reliably be applied to enjoyment profiles because they do not correctly account for the fact that the enjoyment distribution may have multiple peaks. An extreme example of this case is

shown in Figure 16-6. In this case, all viewers have voted that the first and last parts of the video are the best (a bimodal distribution). The standard deviation of this distribution is almost equal to the mean (30s), and the kurtosis is negative, signifying a flat distribution. Therefore, an alternative method for measuring consensus among raters is needed.

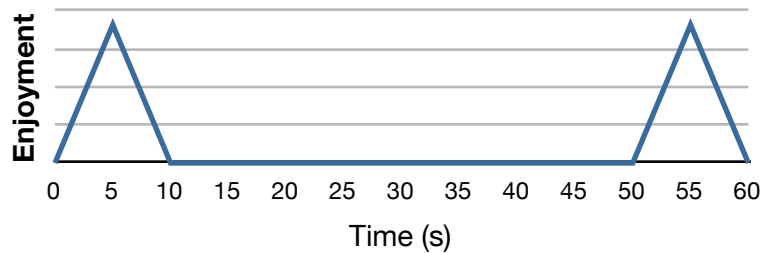


Figure 16-6. Hypothetical enjoyment following a bimodal distribution. This example illustrates why standard measures of spread and peakedness cannot determine consensus in enjoyment profiles. In this case, all viewers have labelled the beginning and end parts of the video as their favorite part.

Clustering is often used to group people based on their ratings of a set of items. Clustering can also be used as a means for determining consensus: by clustering participants based on how they rated each moment of the video, we can tell if there was high consensus (few clusters) or low consensus (many clusters). The momentary ratings are represented as a binary attribute that signifies whether each second in each video was part of each participants' reported favorite sections.

The EM clustering algorithm in the Weka toolkit³⁷ is used to perform the actual clustering (Witten & Frank, 2005). EM was run with its default settings³⁸. EM was used because it does not require the number of clusters to be specified beforehand. Thus, the number of clusters generated by EM can be used as a metric for the degree of consensus. Other common clustering algorithms, such as k-means, require choosing the number of clusters ahead of time.

One consideration that must be addressed with clustering is how to handle participants who did not report any favorite parts in a video. On one hand, these participants signal that no part of the video was enjoyable. On the other hand, including participants who did not vote may interfere with the clustering. For example, a cluster that consists of a participant who voted and a participant who

³⁷ Weka toolkit. <http://www.cs.waikato.ac.nz/ml/weka/>

³⁸ EM -I 100 -N -1 -M 1.0E-6 -S 100, mode: use training set.

didn't does not necessarily indicate that those participants had a high degree of consensus. Thus, participants who did not vote for a video are not included in the clustering.

To determine consensus, five values for each video are reported: the average rating of the video across all participants on a 5-point scale (5 highest), the number of participants who reported favorite sections in the video (rater count), the number of clusters produced by EM, the size of the largest cluster, and the ratio of the largest cluster size to the rater count. The degree of consensus is determined by having fewer, larger clusters, or by having a small rater count. Table 16-4 reports the results of the clustering-based consensus analysis.

Table 16-4. Results of clustering participants to determine consensus on labeling the best parts of videos. Only participants that reported at least one favorite part for a video were included in the clustering; this count is reported as the rater count. Video ratings were made on a 5-point scale (5 highest). The largest cluster ratio is the ratio of the largest cluster size to the rater count.

Video	Mean rating	Rater count	Cluster count	Largest cluster size	Largest cluster ratio	Consensus
Cartoon study						
Emerge	2.05	11	2	9	81.8%	High
Fuggy Fuggy	2.5	12	3	7	58.3%	Moderate
In the Rough	3.45	18	2	15	83.3%	High
Pen Pals	4.1	18	3	8	44.4%	Moderate
Penguin's Christmas	3.6	18	6	9	50.0%	Low
Plumber	3.75	16	3	9	56.3%	Moderate
War Photographer	2.4	9	1	9	100.0%	High
Text vs. Audio study						
Ali G - War	3.45	17	7	6	35.3%	Low
Brothas From the Same Motha	2.1	12	4	6	50.0%	Low
Daughters	3.1	14	2	11	78.6%	Moderate
Gopher Broke	3.75	18	2	15	83.3%	High
Paddy the Pelican	1.65	3	1	3	100.0%	High
Powaqqatsi	2.05	8	1	8	100.0%	High
Tea	1.6	8	1	8	100.0%	High
Tony vs. Paul	2.9	13	3	11	84.6%	Moderate

As a sanity check, the mean video rating was significantly correlated with the rater count ($r = .92, p < .0001$). Thus, more participants reported favorite sections for videos when those videos were better.

To qualitatively understand the amount of consensus in the hand-created enjoyment profiles, we consider the number of clusters and the ratio of the largest cluster size to the number of raters. High consensus was defined as having one cluster, or two clusters where the largest cluster had more than 80% of the raters. Low consensus was defined as having four or more clusters. Moderate consensus was defined as everything in between.

Seven videos stand out as having high consensus among raters. Four of these videos were not highly rated and less than half of participants reported favorite parts for them (rater count < 10). These videos were “War Photographer” (9 raters, 1 cluster), “Paddy the Pelican” (3 raters, 1 cluster), “Powaqqatsi” (8 raters, 1 cluster), and “Tea” (8 raters, 1 cluster). This set also contained videos that were favorably rated (e.g., “Gopher Broke,” 3.75) and poorly rated (e.g., “Emerge,” 2.05).

Participants displayed a moderate amount of consensus in their enjoyment profiles for five videos. These videos in this set had cluster counts of 2 or 3, and this set included videos that were highly rated (e.g., “Pen Pals,” 4.1) and poorly rated (e.g., “Tony vs. Paul,” 2.9).

Little consensus was present in the three videos having cluster counts of 4 or greater. This set also contained videos that were favorably rated (e.g., “Penguin’s Christmas,” 3.6) and poorly rated (e.g., “Brothas From the Same Motha,” 2.1).

16.4.4. DISCUSSION

The Best Part labeling study addressed several questions about enjoyment profiles: what do they look like, do they show a clear differentiation between enjoyable and less-enjoyable parts of a video, do chat-extracted profiles accurately capture the parts of a video people most enjoy, and how much consensus was present in hand-created profiles? These questions were important because if chat-extracted profiles did not show clear patterns of enjoyment, or if they did not match any sort of reality, their utility would be questionable. In addition, if there was little consensus for which parts of a video are most enjoyable, the task of inferring enjoyment profiles

would need to be reconstructed to account for differences in individuals' preferences for and enjoyment of videos.

Examples of chat-extracted profiles are shown in Figures 16-2, 16-3, and 16-5. Each of these figures show clear patterns of enjoyment versus non-enjoyment. Therefore, the chat-extracted profiles can possibly be representative of actual enjoyment.

To measure whether chat-extracted profiles were actually representative of actual enjoyment, they were compared to profiles constructed by human raters. Overall, the amount of agreement between hand-created enjoyment profiles and chat-extracted enjoyment profiles was variable and dependent on the resolution of the profile. For a high resolution (5 seconds), the percentage agreement between hand-created and chat-extracted profiles ranged from about 35% to 76%. In this case, three videos had agreements between 50%-60%, three videos had agreements between 60%-70%, and one video had an agreement greater than 70%. These results improved when a lower resolution was used (15 seconds). In this case, the percentage agreement between hand-created and chat-extracted profiles ranged from about 46% to 83%. Six videos had agreements between 50%-60%, four had agreements between 60%-70%, two had agreements between 70%-80%, and one had an agreement greater than 80%.

Overall, the agreement results show that it is realistic to construct enjoyment profiles from chat data. For some videos, the chat-extracted profiles had a high agreement with hand-created profiles. For other videos, the agreement was less. One explanation for this result is that low-agreement videos may have had a high amount of non-video related laughter ("noisy" laughter). With more data, the signal to noise ratio increases, thus increasing agreement. This effect was seen between the Cartoon study videos (low data) and Text vs. Audio study videos (high data). Agreement was higher for the videos with more data.

However, all of these results assume that the hand-created enjoyment profiles constituted a "gold standard". Although we may judge an agreement of 50% as being somewhat low, we questioned whether the hand-created profiles were even representative of actual enjoyment. The second analysis in this study considered the degree to which raters agreed with each other in their selection of the best parts of the videos. We found that the raters exhibited varying levels of consensus, independent of the quality of the video. For seven videos, raters agreed with each

other on the locations of the best parts. For four of these videos, consensus emerged simply from a universal dislike of the video, as raters either did not report any favorite parts or reported only very few. Other videos had either moderate or low degrees of consensus, suggesting that when people did not universally dislike a video, their tastes and preferences for what they did like still varied.

The conclusion from this analysis is that enjoyment profiles ought not be universally constructed for all viewers of a video; rather, more accurate profiles may be constructed by clustering users based on their individual tastes and computing profiles for each cluster. For example, consider how this scheme might work for an online video site such as Netflix. Assuming that Netflix collected momentary ratings of enjoyment³⁹, they could display a set of the best parts for each video as rated by other viewers with similar tastes. For example, a fan of science fiction films could see what other science fiction fans thought of the worm scenes in “Dune,” filtering out the opinions of non-science fiction fans who may view those scenes differently.

Finally, there is one comment to be made on the resolution of enjoyment profiles. Resolution is an important feature to consider when computing an enjoyment profile. Although higher resolutions provided finer-grained detail, they also introduced more variance in the profile, making it difficult to interpret the overall trends. This tradeoff between interpretability and resolution can be presented in another way: does our interpretation of what constitutes a high point change when we think it is between 30 and 35 seconds versus when it is between 30 and 45 seconds? This question can only be answered by those people who wish to create and use enjoyment profiles, as it depends on their intended use and needs for precision. This chapter merely expresses the importance of resolution and shows the consequences of using a high versus a low resolution.

16.5. GENERAL DISCUSSION

Laughter is an important part of the media consumption experience. Early studies of group laughter had participants listen to two humorous recordings, one with the

³⁹ This capability can be implemented with little effort. Netflix supports viewing movies on the Xbox 360 simultaneously with other viewers. Further, these viewers can chat with each other using the Xbox’s voice chat feature. Privacy considerations aside, it would not be difficult to add a laughter classifier to each audio stream to build an enjoyment profile similar to those examined in this chapter.

dubbed laughter of a group of prior viewers, and one without (Smyth & Fuller, 1972; Fuller & Sheehy-Skeffington, 1974). In both of these studies, participants laughed more frequently and for longer in the presence of dubbed laughter. In the second study, participants also smiled more and rated the material higher in the presence of dubbed laughter (Fuller & Sheehy-Skeffington, 1974). A follow-up study by Platow et al. (2005) manipulated participants' beliefs as to *who* was laughing: an in-group with whom they identified (members of their university), or an out-group with whom they did not identify. They found that participants laughed and smiled more, laughed longer, and rated humorous material more favorably, when they believed that the laughter came from the in-group. Therefore, not only does laughter indicate moments of enjoyment, it can also be used to enhance others' enjoyment as well.

The moment-by-moment profiles of enjoyment can be used to generate laugh tracks for online videos, automatically, by aggregating laughter across viewers. To follow on the results of the study by Platow et al. (2005), this laughter can be aggregated and disseminated at different levels: across all viewers in an online community (providing an identity connection), across viewers with similar tastes or interests (also providing an identity connection), or across one's social network (providing a bond connection). Aggregating laughter among one's friends may increase one's confidence in the genuineness of the laughter – that the laughter was truly a response to humorous or entertaining material, as opposed to being an artificial or canned response. Genuine laughter is desirable because once laughter is construed as artificial, its effect on audiences is weakened (Lawson, Downing, & Cetola, 1998).

Enjoyment profiles have many other uses as well. For video search engines, enjoyment profiles can be used to generate thumbnails for videos from their most-enjoyed parts. These scenes may help increase peoples' recall when searching for a particular video, because the most-enjoyed scenes may be more recognizable or recalled quicker than boring scenes. For users engaged in a browsing task, the best scenes may also be the most enticing, helping the user decide if he or she wants to watch a particular video.

Video summarization engines can also benefit from having enjoyment profiles. These summarizers often segment a long video into its component clips (e.g., Wactlar, 2000) by detecting changes in the scene (e.g., Huang & Liao, 2001). Using enjoyment profiles, the relative importance of each scene can be computed, and thus a summary of a video can be built from its most-enjoyed scenes. This type of

summarization has many applications, including automatically creating sports highlight reels, reordering news stories based on viewer interest, creating interesting remixes and mash-ups from a large set of videos, or even creating laugh tracks for videos so viewers don't feel they are watching in isolation.

In the world of television and movies, test audiences and focus groups are used to determine how viewers will react to the show or movie (Eliashberg & Sawhney, 1994). In cases where an audience fails to react, such as when a joke isn't funny, or an ending is unsatisfying, writers, producers, and directors can rewrite a script to provide a more enjoyable or more satisfying experience. To improve the accuracy of feedback, as well as provide moment-by-moment evaluations, physical devices have been used to quantify an audiences' reactions. Many of these devices were based on the Lazarsfeld-Stanton Program Analyzer (Fiske & Handel, 1947). This device consisted of two buttons – red and green – held in opposite hands. Audience members were instructed to push the green button during periods of approval, the red button during periods of disapproval, and not press any button when feeling indifferent. Since the 1940s, numerous variations on this device have been developed and used including knobs and joysticks (Millard, 1992).

One limitation of these devices is that they require a conscious effort on the part of the viewer: they must continuously evaluate their own level of enjoyment or approval and turn a knob, push a button, or move a joystick when that level changes. This evaluation method is completely at odds with the activity it is meant to measure. Viewers often watch video content to lose themselves and forget about their immediate situation (Finn & Gorr, 1988). But because this evaluation process requires viewers to actively and continuously think about and evaluate their enjoyment, they may fail to register their enjoyment with the system when they are distracted by the video. This distraction adds error into the enjoyment measure.

When video content is consumed online in a collaborative context, it becomes possible to automatically generate profiles of enjoyment through the chat exchanges that occur. Since chatting with others still requires viewers to actively engage with something other than the video content, they may also fail to register their momentary enjoyment when they are distracted by the video. However, creating enjoyment profiles from collaborative watching has one important benefit compared to its real world analog: scale. Momentary enjoyment measures can be collected from far more viewers in an online space than in the real-world. Further, collecting

this information unobtrusively from online viewers is cheaper than collecting it obtrusively from real-world viewers – online viewers provide their momentary enjoyment as a by-product of social interaction (from which they derive value), whereas real-world viewers are often paid for their time. The only remaining question is then, are enjoyment profiles constructed from online sources any good?

The results presented in this chapter suggest that enjoyment profiles extracted from chat data are useful because they show clear patterns of enjoyment and non-enjoyment, even in the presence of non-video related laughter. They also seem to be accurate, at least for some videos, because they tend to agree with enjoyment profiles constructed by human raters. However, hand-created profiles are not perfectly consistent between viewers – different viewers liked different parts of each video. Therefore, enjoyment profiles should be constructed for viewers having similar tastes. This way, trends can be interpreted with respect to the type of viewer watching the video. For example, a viewer who enjoys schtick may overwhelmingly enjoy a particular joke that another viewer may find only slightly amusing.

This chapter demonstrates that enjoyment profiles can be learned from chat data. This conclusion both reinforces and extends the results found by Miyamori, Nakamura, and Tanaka (2005). Although their study showed that chat data could be mined for expressions of “enjoyment” and “disappointment” to create moment-by-moment profiles of enjoyment, their evaluation only used three human raters. This chapter presented a more comprehensive study that found that the chat-extracted enjoyment profiles did tend to match hand-constructed enjoyment profiles, and that accuracy increased as more data was collected. The goal of this work was not to perfectly match the hand-constructed profiles, especially since several of those profiles exhibited low agreement consensus. Instead, this chapter demonstrates that it the enjoyment profiles extracted from chat did exhibit clear patterns over time, and that they were at least somewhat representative of viewers’ enjoyment. Future work in this area is needed to consider how linguistic features other than laughter indicate other types of emotional reactions, such as surprise, shock, or fear, to video content over time.

16.6. SUMMARY AND CONCLUSIONS

- The finding from Chapter 15 that laughter is indicative of enjoyment suggested that a moment-by-moment profile of enjoyment could be constructed for a video by aggregating laughter in chat across all viewers of a video.
- Enjoyment profiles were constructed for the videos in the Cartoon and Text vs. Audio studies. They showed clear patterns of enjoyment and non-enjoyment.
- The most enjoyed parts (high points) of a video can be determined by locating the points where enjoyment is highest. The resolution of the profiles is an important factor as it affects the interpretation of the profile and the locations of its high points.
- The Best Part Labeling study determined that the chat-extracted enjoyment profiles tended to agree with the enjoyment profiles constructed by human raters.
- Agreement between chat-extracted and hand-created enjoyment profiles ranged from 36% to 76% for a 5-second resolution, and from 46% to 83% for a 15-second resolution. Agreement was higher for cases where more data was available. These results demonstrate that it is feasible to automatically construct enjoyment profiles from chat data in cases where enough data are available.
- Hand-created enjoyment profiles exhibited varying degrees of consensus among raters for where the most enjoyable parts were located. High consensus was present for 7 of 15 videos, a moderate amount of consensus was present for 5 videos, and a low amount of consensus was present for 3 videos. These individual differences in enjoyment suggest that profiles may need to be created only among viewers with similar tastes and preferences for videos.

Part V: Conclusions

The Internet has revolutionized the way in which we consume video content.

In Part V, I discuss future research directions that will further our understanding of collaborative online video watching and help us design experiences for all of the different types of video content offered online.

17.

LIMITATIONS AND FUTURE WORK

This dissertation scratches the surface of a new and interesting type of online activity: chatting while watching videos. In this chapter, I discuss several important limits to the generalizability of the findings in this dissertation, and discuss how future work can overcome those limitations.

17.1. ALTERNATIVE VIDEO CONTENT

This dissertation studied collaborative online video with an emphasis on content that was entertaining or funny. Alternate forms of content should be considered as well, including content that is meant to be informative (i.e., politically-themed or newsworthy), content that is meant to have an emotional impact (i.e., a drama or documentary), and content that is meant to be controversial or raise awareness of an important issue (i.e., anything by Michael Moore). It is possible that a chat feature is less appropriate or more distracting when watching other types of content. In these cases, a structured experience may be necessary to mitigate the distracting effects of chat. For example, it is worse when a viewer misses an important topic in an online lecture because they were chatting than when they miss the punch line to a joke in a comedy routine. Intermissions may be required to help viewers focus on important parts of the lecture.

To mitigate this limitation, the study of the Social Video application in Chapter 12 did employ several informative videos, including lectures. These videos generally required more cognitive resources to process because they required thinking and interpretation. Some of the participants who watched these videos did use the chat feature while watching. Therefore, it may be possible for people to manage their

attention while watching informative content, although participants in this situation weren't quizzed at the end to test their recall. Thus, a more thorough examination of the distractive effects of chat for different types of video content is required.

Specific genres of videos that may have a higher information-processing demand on viewers include:

- Sporting events,
- News programs,
- Talk shows,
- Documentaries,
- Films,
- Political speeches,
- Debates,
- Lectures, and
- Surveillance video (e.g., traffic cameras or web cams in public spaces).

In addition to examining the distractive effects of chat on these genres, the impact of chat media should be considered as well. For example, voice chat may work well for sporting events, during which viewers often express excitement and frustration out loud during the course of a game. Contrarily, political debates typically require a greater amount of attention in order to learn about the candidates and their positions, and voice chat may interfere with this process.

Another facet of video content is the length of the video. Many of the studies in this dissertation used short, 3-6 minute long videos. However, for many of the genres listed above, their content is much longer. The MovieLens study did employ longer content (2-3 hours), although participants tended to diminish their chat activity as they became engrossed in the movie. More work is needed to determine how a chat feature should be best utilized for longer content. Should a movie be paused to give people time to chat without being distracted? Should the 'chat' consist of a synchronous, back-and-forth chat, or should it operate more like the Facebook Live Stream box in which people post short messages? Should a moderator be used to prompt people to talk when they are quiet? How do these decisions affect people's feelings of distraction?

Finally, more work is needed to determine peoples' interest in chatting while watching videos with differing amounts of a priori emotional significance. For example, a person who is anticipating the release of a new movie may not want to chat at all during their first viewing because they want to partake in a completely engrossing experience. Or, a person who is anticipating the broadcast of a political debate may want to chat with others as they will only have one opportunity to do so. Contrarily, a person who feels indifferent toward a movie may prefer to chat while watching as it gives them something else to do during the uninteresting parts.

17.2. COLLECT MORE REAL-WORLD DATA

This dissertation makes heavy use of laboratory studies to draw conclusions about a highly popular online activity. As discussed earlier, the laboratory studies allow for a high degree of control of the confounding factors that exist in the real-world. This control comes at a loss of generality. Participants in the lab studies may have been unduly influenced by the laboratory context to use the chat features provided to them in ways that may not reflect how they would actually use those features in a natural setting. In addition, most participants were college students, further limiting the generalizability of the findings.

Perhaps the most suspect finding in the laboratory studies is that strangers enjoyed chatting with each other while watching videos (Chapter 8). Studies of Facebook usage by Lampe, Ellison, and Steinfield (2006) and Joinson (2008) suggest that people are not very interested in meeting new people on Facebook. Rather, they use it to remain connected to their existing group of friends. The survey in Chapter 11 also showed that people were most interested in chatting with their friends while watching videos, and were not much interested in chatting with strangers. Thus, given that there may be resistance toward chatting with strangers while watching videos, and given that chatting with strangers is desirable (the sociability argument), how can people be encouraged to do so?

The designs in Chapter 12 for promoting "stranger encounters" – visualizations and summaries of the activity of strangers – are a good starting point. They capitalize on the fact that a viewer does not need to directly interact with others in order to feel their presence. In this way, the visualizations and summaries may promote the formation of 'familiar stranger' relationships between viewers. This type of

relationship exists when people who are strangers to each other come to recognize each other from repeated participation in some activity or event, such as riding a bus every day (Milgram, 1977). In this relationship, there is no direct interaction; once there is interaction, a stronger relationship can be formed. Instead, the presence (or conspicuous absence) of the other is enough to maintain the relationship. Visualizations such as the large audience proxy can be equipped to support these types of relationships by selecting interesting audience members and highlighting them in the interface, for some definition of interesting. For example, a person recommendation system can be used to find viewers with similar likes or interests in their profile, and these recommendations can be displayed in the audience proxy. Future research should be conducted to understand how such recommendations help viewers feel more connected to the audience and develop relationships with other audience members.

17.3. ADDITIONAL BEHAVIORAL MEASURES

The studies in this dissertation employed several self-reported and behavioral measures. Enjoyment was measured by having participants rate their enjoyment or rate the quality of the videos on a scale. Distraction was measured by asking participants to rate their feelings of distraction on a scale and by asking them to recall what they had watched the videos. The measure of engagement – eating or not eating pretzels – seemed to make sense theoretically when it was developed, although in practice the social norms of not chewing loudly precluded its usefulness.

In general, behavioral measures are more desirable because they are more accurate. Chapters 15 and 16 explored behavioral measures of enjoyment by considering when and how much participants laughed while watching the videos. Other behavioral measures of enjoyment, engagement, and distraction may exist as well. For example, physiological measures of the sympathetic nervous system, including heart rate, skin temperature, blood pressure, and respiratory rate, as well as measures of galvanic skin response and pupil dilation, are commonly used to measure emotional arousal and/or stress (e.g., Zellars et al., 2009; Bradley et al., 2008; McCleary, 1950). Measures such as these might be adopted to provide more sensitive and momentary measures of engagement and/or enjoyment. In addition, eye tracking can be used to understand how much time viewers spend looking at videos and looking at chat. However, one caveat to these measures is that they are

somewhat invasive and distracting themselves (imagine trying to type chat messages to your friends while wearing a pulse meter!). Thus, their utility for more accurately measuring enjoyment and distraction during a collaborative watching session should be evaluated.

17.4. VISUALIZATION DISTRACTION

The visualizations discussed in Chapter 11 provide continuously-updating displays of chat activity in a collaborative online video system. These displays are another visual source of information that contend for a viewer's attention. Now, they must multitask between the video, the chat of their group, the summary of chat from the rest of the audience, and the dynamic representation of that audience. These additional displays may further increase levels of distraction while watching a video. Future work should be conducted to measure the degree to which viewers are distracted from these additional information sources. This work should be conducted in a simulated environment, in which the amount of chat and the rate at which the visualizations update their information are controlled.

In addition, the distraction of interactive visualizations should be considered. We have seen in this dissertation that viewers are able to manage their attention between chatting and watching videos. If audience representations and chat summaries allow (or require) interactive exploration, will viewers be able to manage their attention in this case, or will the requirements of interaction be overly distracting? Future work is needed to address this question.

17.5. SOCIAL DASHBOARDS

The observation in Chapter 11 that the audience proxy could be used to represent either historical or current activity motivates the creation of an interactive social dashboard that helps people understand their mutual activity with their friends. This dashboard can be used as an additional mechanism for promoting interactions among friends by revealing people with whom one has not recently communicated. It can also raise the visibility of friends-of-friends and strangers by showing people who are two or three steps away in one's social network.

The concept of a social dashboard has been explored in several domains. Ducheneaut et al. (2007) developed a social dashboard for players of the World of Warcraft game. This dashboard helped players visualize the composition of their guild and showed them areas in which they could improve. In their study of guilds, Ducheneaut et al. found that guilds with a diverse spread of players at different levels and of different classes lasted longer than guilds with less diversity. Their social dashboard made this diversity apparent by showing the areas in which guilds should focus their recruitment efforts.

Suh et al. (2008) developed a dashboard for Wikipedia that showed the edit history for each article. This tool was designed to increase the transparency of editing activity and the accountability of editors. An early evaluation suggested that the increased visibility of editing history improved the interpretation, communication, and trustworthiness of Wikipedia articles.

Research on the effect that social dashboards have on their ability to help people maintain ties is lacking. The studies by Ducheneaut et al. (2007) and Suh et al. (2008) found that users liked their respective dashboards, but not that usage of the dashboards was quantitatively associated with increases in guild lifetimes or article quality. In addition, these dashboards were not focused on helping people manage their *social* relationships so much as they were focused on improving the quality of the guild or the quality of the articles. Thus, there remains an open design problem for creating a truly social dashboard – one that helps people create and/or maintain ties in their social network – as well as quantitatively evaluating whether the dashboard has a positive, longitudinal impact on social capital.

17.6. IMPROVED TEXT MINING

The text mining algorithms discussed in Part IV are all somewhat simplistic. They treat the text corpus as a bag of words and simply count and weight terms accordingly. Other algorithms from machine learning and computational linguistics can be employed to infer more accurate or meaningful data from “messy” text chats, including:

- Clustering chat messages into topic areas using latent dirichlet allocation (Blei, Ng, & Jordan, 2003), thread detection (Shen et al., 2006) or segmentation (Utiyama & Isahara, 2001),

- Hidden Markov models to explore relationships between messages, senders, and/or topics (Rabiner, 1989),
- Conditional random fields to segment messages and classify them into topic areas (Lafferty, McCallum, & Pereira, 2001),
- Keystroke-level models to determine the excitement level present in typed messages; prior work in this domain has shown that individuals can be differentiated by their typing styles (Bryan & Harter, 1897; Joyce & Gupta, 1990), and that increases in arousal can alter one's typing style (Henderson et al., 1998).

This dissertation shows that interesting and useful information *can* be learned from messy chat data. Future work is needed to further improve the quality of the information that we *do* infer.

17.7. SUMMARY AND CONCLUSIONS

- Alternative video genres should be studied to understand how they impact viewers' distraction from chat. These genres include news and political broadcasts, sporting events, and documentaries.
- More real-world data is needed to determine the extent to which the design of a collaborative online video site can encourage interactions among strangers who may not be interested in interacting.
- Physiological measures such as pulse, galvanic skin response, and eye-tracking can provide more accurate measures of enjoyment, engagement, and distraction. They may also interfere with the task of watching and chatting, and thus their utility should be evaluated.
- Visualizations of a large audience and their chat messages may be additionally distracting. Controlled laboratory studies are required to measure this additional distraction and to determine if the positive effects of the visualizations (e.g., feeling connected to the audience) outweigh the negative effects (e.g., being additionally distracted).
- Social dashboards display information about the activities of others online. They are a helpful tool for increasing the quality of contributions and participation in online communities such as Wikipedia and World of

Warcraft. Further research is needed to determine their effectiveness in helping people create and/or manage relationships with others.

- More sophisticated machine learning and computational linguistics algorithms may be able to infer more accurate information from “messy” chat data.

18.

CONCLUSIONS

Watching videos is one of the most popular applications on the Internet today. Like the other major technology used for distributing and consuming video content – the television – watching videos online is capable of supporting social interactions *before, during, and after* the act of consumption. Unlike the television, online video places no requirement on physical co-locality: viewers can watch videos with others no matter where they are located, as long as they have an Internet connection.

This freedom from the constraints of physical reality opens up tremendous possibilities for improving the state of social interactions around video content. Putnam (1995, 2001) has argued, very convincingly, that social capital in America has been on the decline, and that television is one of the causes. Watching television is often done in solitary (Lee & Lee, 1995), and precludes activities that promote social interaction and building social capital, such as spending time at a “third place” like a bowling alley, a bar, or a coffee shop (Oldenburg, 1999). Online video holds the promise of creating new “third places” online, by combining videos (the online equivalent of a cup of coffee, a stein of beer, or a set of bowling pins) with social interaction features (the online equivalent of a conversation). In this way, online video sites become conduits through which remote viewers watch and interact with each other, enabling the building and maintenance of social capital.

But do people really want to have social interactions while watching a video? Isn't talking while watching television – essentially the main activity examined in this dissertation – rude and distracting? Why should it be any different online?

In fact, this dissertation demonstrates that it is different. We did see interference between the activity of chatting with others and the activity of watching videos. But, for the videos studied, the magnitude of this interference was small. People found enough value and enjoyment in chatting with others while watching that they kept using the chat feature, even when they were given breaks to chat without interference, and even when they were in control their own video playback and could pause to chat.

For many viewers, watching television is an escapist activity that helps them relax and escape from ordinary day-to-day cares, at least for a little while (Lee & Lee, 1995). In each study presented in this dissertation, there consistently were a small set of participants who expressed disinterest in the chat feature. This disinterest was sometimes expressed behaviorally, such as when participants did not use the chat features provided, or did not use them much. In other cases, this disinterest was directly reported, such as when participants said that they simply would have preferred not to chat. In a real-world setting, viewers cannot be 'forced' to chat; lurking is, and will always, be a reality of online communities. Indeed, Nonnecke and Preece (2000) argue that lurking may in fact be 'normal' behavior, and without lurkers, there may not be anyone to read the messages posted by others. Thus, although lurkers may not contribute directly to conversations, features can still be designed for them that provide them an awareness of the conversations of others and help them feel connected to those other viewers. These features include the visual summaries of chat and the social proxy audience representation discussed in Chapter 11. Therefore, watching videos online can still be a social experience, even for those viewers who choose not to actively participate in the conversation.

This dissertation furthers our understanding of the *collaborative* online video experience. It demonstrates that watching collaboratively is enjoyable and leads to momentary gains in feelings of sociability. It demonstrates how to provide this experience in scale, when audiences are too large to fully comprehend. It demonstrates that the increasing use of the Internet to produce, distribute, and consume video content may also provide us with opportunities to rebuild social capital lost to television. Television may be isolating. Online video need not be.

APPENDIX A: POPULAR ONLINE VIDEO SITES AND SYSTEMS

Online video is one of the most popular applications on the Internet today. Many sites and systems have been created for the purpose of delivering video content online. Some of these services incorporate social features, some provide access to different types of specialized content, and some were developed for the purpose of conducting research on the social viewing experience. For historical reference, Table A-1 summarizes some of the currently popular online video sites, shows, and systems.

Table A-1. List of popular online video sites and systems. Sites were selected for inclusion on the basis of popularity, uniqueness of content, social interaction features, or discussion in this dissertation. The year listed is the year in which the site was founded, the video component of the site was launched, or the technology was released. Descriptions of each category are given and are generally applicable to each site listed in that category; additional description is given for each site where appropriate.

Site/Show/System	Year	Description & Notes	URL / Reference
User Generated Content - Upload			
YouTube	2005	<i>similar to category description</i>	youtube.com
Yahoo Video	2006	<i>similar to category description</i>	video.yahoo.com
AOL Video	2006	<i>similar to category description</i>	video.aol.com
Google Video	2006	Videos aggregated from other online video sites	video.google.com
Vimeo	2004	Focus on high-quality video (e.g., HD) and visualizing community activity	vimeo.com
Metacafe	2003	Focus on short-form, entertaining videos	metacafe.com
Dailymotion	2005	Special section of the site devoted to videos for kids	dailymotion.com
CollegeHumor	1999	Focus on original comedy videos and articles	collegehumor.com
Break.com	1998	Comedy and humor videos targeted at the male 18-34 demographic	break.com
Viddler	2006	Publish videos for personal use; revenue sharing model for businesses	viddler.com
GotGame	2006	Focus on video games (in-game clips and promotional videos)	gotgame.com
JewTube	2006	Focus on religious videos (Jewish)	jewtube.com
Tangle	2007	Focus on religious videos (Christian)	tangle.com
User Generated Content - Streaming			
		Users stream live video from their computers, game consoles, and mobile devices	

Site/Show/System	Year	Description & Notes	URL / Reference
Justin.TV	2006	Text chat and Twitter integration in video player	justin.tv
UStream.TV	2006	Text chat and Twitter integration in video player	ustream.tv
Livestream	2007	Twitter integration in video player	livestream.com
Kyte.TV	2006	Focus on community-building and monetization	kyte.tv
Social networking sites		Video feature allows users to upload and share home videos	
Facebook	2004	<i>similar to category description</i>	facebook.com
MySpace	2006	Portal for sponsored videos and videos on other sites	vids.myspace.com
Educational / Informative		Video sites providing educational materials and sharing inspired thinking	
TED	2007	Inspirational and informative lectures	ted.com
OpenCourseWare	2008	Educational videos	ocw.mit.edu
News, politics, & current events		Sites and shows focused on keeping viewers informed about news, politics, and current events	
C-SPAN	2005 (podcasts)	Live video stream and video podcasts from the C-SPAN network	c-span.org
CNN/Facebook	2009	Partnership between CNN and Facebook integrates live video and social networks for important political happenings (screenshot in Figure 1-1)	edition.cnn.com/video/fb/facebook.html
MSNBC	2005	Video player allows clips to be queued for sequential playback	tv.msnbc.com
Current.TV	2005	Recent comment summarization on home page	current.tv
LiveLeak	2006	Promotes citizen journalism	liveleak.com
Bill Moyers Journal	2007	Text transcripts of each show make videos more accessible	pbs.org/moyers/journal
The Daily Show	2007	"Wayback randomizer" lets people watch random clips from the show's history	thedailyshow.com
Television & movies		Major networks & studios provide access to their television and movie content online	
ABC	2006	<i>similar to category description</i>	abc.go.com/watch
CBS	2008 (social viewing)	Social viewing with a text chat feature	cbs.com/socialroom
NBC	2008 (P2P)	Uses P2P technology to deliver video	nbc.com
Fox	<i>unclear</i>	<i>similar to category description</i>	fox.com
Hulu	2007	Sponsored by News Corp., provides access to television shows from major networks	hulu.com
Joost	2006	<i>similar to category description</i>	joost.com
Veoh	2004	<i>similar to category description</i>	veoh.com
Netflix	2009 (XBox viewing)	Subscribers can watch streaming videos online; XBox 360 integration provides voice chat feature to friends watching together	netflix.com
Lycos Cinema	2006-2009	Watch full-length movies; shut down in 2009 due to lack of mainstream content	<i>(defunct)</i>
Lostpedia	2005	Members watch Lost on television while posting to forums or participating in a live chat	lostpedia.com
FirstShowing	2006	Members attend movie premiers together	firstshowing.net

Site/Show/System	Year	Description & Notes	URL / Reference
Cinema Blend	2004	Members have live chats during important events (e.g., the Oscars)	
Sports			
JumpTV	2004	Subscription packages provide access to world-wide sporting events	jumptv.com
ESPN	<i>unclear</i>	Sports highlights and shows	espn.go.com/video/
Pornography			
YouPorn	2006	Users can watch streamed clips and upload their own videos	youporn.com
RedTube	<i>unclear</i>	#3 ranked adult site on alexa.com	redtube.com
Streammate	2003	Adult webcams with chat, pay-per-view system for private shows.	streammate.com
Internet TV shows & podcasts		Video content created specifically for Internet audiences	
Channel Frederator	2005	Features viewer-submitted cartoons, cartoons for kids, and vintage public domain cartoons	channelfrederator.com
PurePwnage	2004	Internet TV show focused on gamer culture	purepwnage.com
Rocketboom	2004	Comedic video blog / newscast	rocketboom.com
Red vs. Blue	2003	Machinima series based on the Halo video games	redvsblue.com
The Guild	2007	Online sitcom based on World of Warcraft players	watchtheguild.com
The Scene	2006	Miniseries about film piracy	welcometothescene.com
Other communities offering video			
Gaia Online	2007 (Gaia Cinemas)	Gaia Cinemas allows users to watch videos together in a 2D avatar environment	gaiaonline.com
Second Life	2003	Video can be embedded into the 3D environment enabling virtual movie theaters	secondlife.com
Social TV research systems		Systems developed specially for research in social and interactive television	
2BeOn	2001	Integrated IM, voice chat, and video conferencing on the television	(Abreu, Almeida, & Branco, 2001)
Reality IM	2003	Chat bot provides real-time information about television programs; text chat with friends while watching	(Chuah, 2003)
AmigoTV	2004	Shows avatars of friends on the television; includes voice chat feature	(Coppens, Trappeniers, & Godon, 2004)
Media Center Buddies	2004	Combines IM with television	(Regan & Todd, 2004)
Telebuddies	2006	Audience interaction through quiz games and trivia contests	(Luyten et al., 2006)
Cha.TV	2006	Creates ad-hoc communities of television viewers using audio fingerprinting to identify viewers watching the same programs	(Fink, Covell & Baluja, 2006)
Social TV (STV1, STV2, STV3)	2008	Connects living rooms of friends and family using open microphones	(Harboe et al., 2008a; Harboe et al., 2008b)
Social Video	2009	The collaborative online video system created for this dissertation	Chapter 12; apps.facebook.com/social_video
Zync	2007	Plugin for Yahoo Messenger lets friends watch YouTube videos together while chatting	(Liu et al., 2007)

Site/Show/System	Year	Description & Notes	URL / Reference
Peer-to-peer systems (P2P)		Applications that allow users to publish and view live streaming video over the Internet using peer-to-peer technologies	
End System Multicast (ESM)	1999-2007	P2P video streaming system with text chat for viewers; technology commercialized in 2007	esm.cs.cmu.edu
CoolStreaming	2004-2005	Technology based on BitTorrent	<i>(defunct)</i>
PPLive	2004	Focus on Chinese television content	pplive.com/en
TVU	2005	Live TV from around the world; monetization platform for content owners	tvunetworks.com
Sopcast	2006	<i>similar to category description</i>	sopcast.com

APPENDIX B: SCALES AND MEASURES

This appendix contains most of the scales and measures used in the studies in Part II. Cronbach's α was computed for each multi-item scale as a measure of reliability – the degree to which each item in the scale measured the same underlying construct (Cronbach, 1951). Scales with α values of .7 or greater are generally considered to be reliable.

The specific presentation of these scales has been changed to fit the style of this document. Items marked with an asterisk (*) were reverse coded.

Enjoyment – Cartoon study

$\alpha = .93$

Please rate your degree of agreement or disagreement with the following statements.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
I had fun watching the cartoons	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The cartoons were entertaining	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Chat enjoyment – Cartoon study

$\alpha = .89$

Please rate your degree of agreement or disagreement with the following statements.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
I had fun chatting while watching the cartoons	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I enjoyed reading what other people said in the chat	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I enjoyed chatting with other people	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Chat enjoyment – Text vs. Audio study **$\alpha = .78$**

Please rate your degree of agreement or disagreement with the following statements.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
I enjoyed talking with the people in my group while watching the videos	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I would have preferred to watch the videos alone*	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The chat added to my understanding of the videos	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The chat added to my enjoyment of the videos	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Liking – Cartoon & Text vs. Audio studies **$\alpha = .81$ (C), $\alpha = .85$ (TA)**



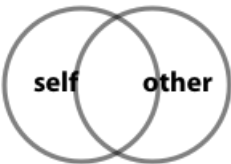
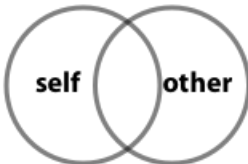
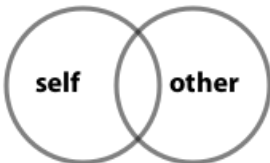
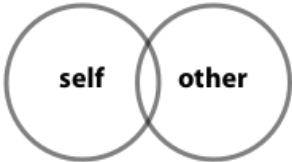
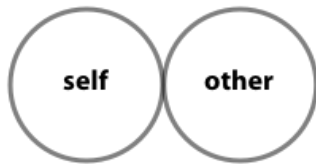
Please describe the other participants in the study. Please answer honestly. Your answers will be kept confidential and will not be seen by the other participants.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
They were friendly	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I liked them	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
If I had to watch more cartoons, I would want to watch them with this same group	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I felt like there was a feeling of togetherness	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

The questions on the liking scale were based on questions from the Work Group Cohesion scale in Price and Mueller (1986).

Closeness scale – Cartoon & Text vs. Audio studies (based on Aron et al., 1991)

During the study, how close did you feel to participant _____ ?

- 
- 
- 
- 
- 
- 
- 

Distraction – Cartoon & Text vs. Audio studies

How distracted were you by the chat during the videos?

Not distracted at all						Very distracted
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Study Enjoyment – Text vs. Audio study

How would you rate the experience of participating in this study?

Very boring						Very fun
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Media Comfort – Text vs. Audio study

$\alpha = .71$

With regard to the {text, audio} chat, please rate your degree of agreement or disagreement with the following statements.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
I found it easy to understand what the other people were saying	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I felt comfortable {typing, talking} while watching the videos	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Memory – Chat Distraction & Text vs. Audio studies

Participants were asked questions about the particular content of each video, including things seen in the visual channel and things heard in the audio channel. Each question was multiple choice, with an option for “I do not recall.” The number of incorrectly answered questions was counted and used as a measure of distraction. Below are several questions used for videos in the Chat Distraction and Text vs. Audio studies. Correct answers are shown in bold text.

Chat Distraction study (video questions)

Brothas From the Same Motha: What did they replace the pen with on the logo? Ray gun, **Futuristic missile launcher**, Space time disrupter, Alien grenade launcher, I do not recall.

Daughters: Where were the terrorists planning on detonating the bomb? Washington, DC, **New York City**, Los Angeles, Chicago, I do not recall.

Tony vs. Paul: What did Paul call Tony in the letter? Stupid, Moron, **Jerk**, Idiot, I do not recall.

Chat Distraction study (chat questions)

Ali G – War: Alex talked about her outfits in the 4th grade. What color did she say she wore? Red, **yellow**, green, black, I do not recall.

Brothas From the Same Motha: Who did not own a cabbage patch kid doll when they were young? Sara, Elaine, **Ted**, Alex, I do not recall.

Tea: What did the stones in the video remind Alex of? A snake, A crocodile, **A salamander**, A gecko, I do not recall.

Text vs. Audio study

Ali G – War: What did Scowcroft say was the “bestest tactic” in war? Aggressiveness, Bigger guns, **Surprise**, Quickness, I do not recall.

Gopher Broke: What animal did the gopher get squished by at the end? **Cow**, Donkey, Horse, Pig, I do not recall.

Paddy the Pelican: Why did the boat not start? **Out of gas**, Broken rudder, Filled with water, Missing oars, I do not recall.

Tea: What kind of tea was poisoned? Black tea, Earl Gray, Iced tea, **Lemon tea**, I do not recall.

APPENDIX C: LAUGHTER EXTENSION TO LIWC

The Linguistic Inquiry and Word Count (LIWC) classifier defines classes for many different kinds of language (Pennebaker, Francis, & Booth, 2001). For example, LIWC defines classes for pronouns (“I,” “our,” “we”), affect (“happy,” “ugly,” “bitter”), cognitive processes (“cause,” “know”) and even leisure activities (“house,” “TV,” “music”). One limitation to the LIWC dictionary is that it does not define a class for laughter. In the studies in this dissertation, participants frequently emitted textual representations of their laughter, such as “haha” and “hehe.”

To perform a linguistic analysis of laughter, I supplemented the standard LIWC dictionary with regular expressions that classify laughter. These regular expressions are not comprehensive over the entire space of how one might laugh over an Internet text channel; rather, they were developed for the specific data sets collected in my studies. The regular expressions I used for classifying laughter are given in Table C-1. Note that these expressions include “jaja,” the common way of expressing laughter in Spanish.

Table C-1. Regular expressions for classifying textual laughter.

“haha” Expressions	“hehe” Expressions	Other Expressions
ha	heh	lol
hah	hee	lolo(.*)
hahh	hehe(.*)	lmao
hahaha	hheh(.*)	lmfao
haa(.*)	hehh(.*)	rotfl
haha(.*)		rofl
hhaha(.*)		jaja(.*)
ahaha(.*)		jajja(.*)
ahhaha(.*)		
haah		

REFERENCES

- Abreu, J., Alemida, P., and Branco, V. (2001). 2BeOn - interactive television supporting interpersonal communication. In J. A. Jorge, N. Correia, H. Jones, and M. B. Kamegai (Eds.), *Proceedings of the Sixth Eurographics Workshop on Multimedia 2001* (pp. 199-208). New York: Springer Verlag.
- Abreu, J. F., Almeida, P., Pinto, R., and Nobre, V. (2009). Implementation of social features over regular IPTV STB. In *Proceedings of EuroITV 2009*, ACM, New York, NY, 29-32.
- Amento, B., Harrison, C., Nathan, M., and Terveen, L. (2009). Asynchronous communication: Fostering social interaction with CollaboraTV. In P. Cesar, D. Geerts and K. Chorianopoulos (Eds.), *Social Interactive Television: Immersive Shared Experiences and Perspectives*. Hershey, PA: IGI Global.
- Arguello, J., Butler, B., Joyce, E., Kraut, R., Ling, K. S., Rosé, C., and Wang, X. (2006). Talk to me: Foundations for successful individual-group interactions in online communities. In *Proceedings of CHI 2006*, ACM, New York, NY, 959-968.
- Aron, A., Aron, E. N., Tudor, M., and Nelson, G. (1991). Close relationships as including other in self. *Journal of Personality and Social Psychology*, 60 (2), 241-253.
- Baecker, R. (2003). A principled design for scalable internet visual communications with rich media, interactivity, and structured archives. In *Proceedings of the 2003 Conference of the Centre For Advanced Studies on Collaborative Research*, IBM Centre for Advanced Studies Conference, IBM Press, 16-29.
- Baecker, R., Baran, M., Birnholtz, J., Chan, C., Laszlo, J., Rankin, K., Schick, R., and Wolf, P. (2006). Enhancing interactivity in webcasts with VoIP. In *CHI'06 Extended Abstracts on Human Factors in Computing Systems*, ACM, New York, NY, 235-238.

- Birnholtz, J., Finholt, T. A., Horn, D. B., and Bae, S. J. (2005). Grounding needs: Achieving common ground via lightweight chat in large, distributed, ad-hoc groups. In *Proceedings of CHI 2005*, ACM, New York, NY, 21-30.
- Blei, D. M., Ng A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3, 993-1022.
- Boertjes, E., Klok, J., Niamut, O., and Staal, M. (2009). ConnectTV: Share the experience. In P. Cesar, D. Geerts and K. Chorianopoulos (Eds.), *Social Interactive Television: Immersive Shared Experiences and Perspectives*. Hershey, PA: IGI Global.
- Bos, N., Olson, J., Gergle, D., Olson, G., and Wright, Z. (2002). Effects of four computer-mediated communications channels on trust development. In *Proceedings of CHI 2002*, ACM, New York, NY, 135-140.
- Bradley, M. M., Miccoli, L., Escrig, M. A., and Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, 45 (4), 602-607.
- Brosh, E., Levin, A., and Shavitt, Y. (2007). Approximation and heuristic algorithms for minimum-delay application-layer multicast trees. *IEEE/ACM Transactions on Networking*, 15 (2), 473-484.
- Brown, B. and Barkhuus, L. (2006). The television will be revolutionized: Effects of PVRs and filesharing on television watching. In *Proceedings of SIGCHI 2006*, ACM, New York, NY, 663-666.
- Bryan, W. L. and Harter, N. (1897). Studies in the physiology and psychology of the telegraphic language. *Psychological Review*, 4 (1), 143-157.
- Butler, B. (1999). The dynamics of cyberspace: Examining and modeling online social structure. Ph.D. Dissertation, Graduate School of Industrial Administration. Carnegie Mellon University. Pittsburgh, PA.
- Butler, B., Sproull, L., Kiesler, S., and Kraut, R. (2002). Community effort in online groups: Who does the work and why? In S. Weisband and L. Atwater (Eds.), *Leadership at a Distance*. Lawrence Erlbaum.

- Cadiz, J. J., Balachandran, A., Sanocki, E., Gupta, A., Gruden, J., and Jancke, G. (2000). Distance learning through distributed collaborative video viewing. In *Proceedings of CSCW 2000*, ACM, New York, NY, 135-144.
- Carnegie Mellon Institutional Research and Analysis (2007). Quick Facts, Fall 2007. Retrieved March 2008 from http://www.cmu.edu/ira/Quick%20Facts/quickfacts_fall_2007_final_web_version.pdf
- Cartwright, D. and Zander, A. (1953). Group cohesiveness: Introduction. In D. Cartwright and A. Zander (Eds.), *Group Dynamics: Research and Theory*. Evanston, IL: Row Peterson.
- Castro, M., Druschel, P., Kermarrec, A., Nandi, A., Rowstron, A., and Singh, A. (2003). SplitStream: high-bandwidth multicast in cooperative environments. In *Proceedings of SOSP 2003*, ACM, New York, NY, 298-313.
- Chi, E. H. and Mytkowicz, T. (2008). Understanding the efficiency of social tagging systems using information theory. In *Proceedings of HT 2008*, ACM, New York, NY, 81-88.
- Chu, Y., Rao, S., Seshan, S., and Zhang, H. (2001). Enabling conferencing applications on the internet using an overlay multicast architecture. *SIGCOMM Computer Communication Review*, 31 (4), 55-67.
- Chu, Y., Ganjam, A., Ng, T. S. E., Rao, S. G., Sripanidkulchai, K., Zhan, J., and Zhang, H. (2004). Early experience with an Internet broadcast system based on overlay multicast. In *Proceedings of USENIX 2004*.
- Chuah, M. (2003). Reality instant messaging: Injecting a dose of reality into online chat. In *CHI'03 Extended Abstracts on Human Factors in Computing Systems*, ACM, New York, NY, 926-927.
- Clancey, M. (1994). The television audience examined. *Journal of Advertising Research*, 34 (4), special insert.
- Collins, N. L. and Miller, L. C. (1994). Self-disclosure and liking: A meta-analytic review. *Psychological Bulletin*, 116 (3), 457-475.

- comScore. (2007). YouTube continues to lead U.S. online video market with 28 percent market share, according to comScore video metrix. Press release. Retrieved October 2009 from http://www.comscore.com/Press_Events/Press_Releases/2007/11/YouTube_Leads_US_Online_Video_Market
- comScore. (2009). YouTube surpasses 100 million U.S. viewers for the first time. Press release. Retrieved October 2009 from http://www.comscore.com/Press_Events/Press_Releases/2009/3/YouTube_Surpasses_100_Million_US_Viewers
- Connell, J. B., Mendelsohn, G. A., Robins, R. W., and Canny, J. (2001). Effects of communication medium on interpersonal perceptions: Don't hang up on the telephone yet! In *Proceedings of SIGGROUP 2001*, ACM, New York, NY, 117-124.
- Coppens, T., Trappeniers, L., and Godon, M. (2004). AmigoTV: Towards a social TV experience. In J. Masthoff, R. Griffiths, and L. Pemberton (Eds.), *Proceedings from the Second European Conference on Interactive Television*. University of Brighton.
- Cosley, D., Frankowski, D., Kiesler, S., Terveen, L., and Riedl, J. (2005). How oversight improves member-maintained communities. In *Proceedings of CHI 2005*, ACM, New York, NY, 11-20.
- Cozby, P. C. (2004). *Methods in behavioral research* (8th ed.). McGraw-Hill, Columbus, OH, USA.
- Cummings, J., Lee, J., and Kraut, R. E., (2004). Communication technology and friendship: The transition from high school to college. In R. Kraut, M. Brynin, and S. Kiesler (Eds.), *Domesticating Information Technology*. Oxford University Press.
- Daft, R. L. and Lengel, R. H. (1984). Information richness: A new approach to managerial behavior and organization design. *Research in Organizational Behavior*, 6, 191-233.
- Daft, R. L. and Lengel, R. H. (1986). Organizational information requirements, media richness and structural design. *Management Science*, 32 (5), 554-571.

- Dennis, A. R., and Kinney, S. T. (1998). Testing media richness theory in the new media: The effects of cues, feedback, and task equivocality. *Information Systems Research*, 9 (3), 256-274.
- Ding, X., Erickson, T., Kellogg, W. A., Levy, S., Christensen, J. E., Sussman, J., Wolf, T. V., and Bennett, W. E. (2007). An empirical study of the use of visually enhanced VoIP audio conferencing: The case of IEAC. In *Proceedings of CHI 2007*, ACM, New York, NY, 1019-1028.
- Dourish, P. (2006). Re-spacing-place: "Place" and "space" ten years on. In *Proceedings of CSCW 2006*, ACM, New York, NY, 299-308.
- Ducheneaut, N., Yee, N., Nickell, E., and Moore, R. J. (2007). The life and death of online gaming communities: A look at guilds in World of Warcraft. In *Proceedings of SIGCHI 2007*, ACM, New York, NY, 839-848.
- Ducheneaut, N., Moore, R. J., Oehlberg, L., Thornton, J. D., and Nickell, E. (2008). Social TV: Designing for distributed, sociable television watching. *International Journal of Human-Computer Interaction*, 24 (2), 136-154.
- Eliashberg, J. and Sawhney, M. S. (1994). Modeling goes to Hollywood: Predicting individual differences in movie enjoyment. *Management Science*, 40 (9), 1151-1173.
- Ellis, C. A., Gibbs, S. J., and Rein, G. L. (1991). Groupware: some issues and experiences. *Communications of the ACM*, 34 (1), 38-58.
- Ellison, N. B., Steinfield, C., and Lampe, C. (2007). The benefits of Facebook "friends": Social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication*, 12, 1143-1168.
- Erickson, T., Smith, D. N., Kellogg, W. A., Laff, M., Richards, J. T., and Bradner, E. (1999). Socially translucent systems: Social proxies, persistent conversation, and the design of "Babble". In *Proceedings of SIGCHI 1999*, ACM, New York, NY, 72-79.
- Erickson, T. and Kellogg, W. A. (2000). Social translucence: An approach to designing systems that support social processes. *ACM Transactions on Computer-Human Interaction*, 7 (1), 59-83.

- Erickson, T., Huang, W., Danis, C., and Kellogg, W. A. (2004). A social proxy for distributed tasks: Design and evaluation of a working prototype. In *Proceedings of SIGCHI 2004*, ACM, New York, NY, 559-566.
- Erickson, T., Kellogg, W. A., Laff, M., Sussman, J., Wolf, T. V., Halverson, C. A., and Edwards, D. (2006). A persistent chat space for work groups: The design, evaluation and deployment of loops. In *Proceedings of DIS 2006*, ACM, New York, NY, 331-340.
- Fink, M., Covell, M., and Baluja, S. (2006). Social- and interactive-television applications based on real-time ambient audio identification. In G. Doukidis, K. Chorianopoulos, and G. Lekakos (Eds.), *Proceedings of EuroITV 2006*, 138-146.
- Fink, M., Covell, M., and Baluja, S. (2008). Mass personalization: Social and interactive applications using sound-track identification. *Multimed. tools appl.*, 36, 115-132.
- Finn, S. and Gorr, M. B. (1988). Social isolation and social support as correlates of television viewing motivations. *Communication Research*, 15 (2), 135-158.
- Fish, R. S., Kraut, R. E., and Root, R. W. (1992). Evaluating video as a technology for informal communication. In *Proceedings of CHI 1992*, ACM, New York, NY, 37-48.
- Fish, R. S., Kraut, R. E., Root, R. W., and Rice, R. E. (1993). Video as a technology for informal communication. *Communications of the ACM*, 36 (1), 48-61.
- Fiske, M. and Handel, L. (1947). New techniques for studying the effectiveness of films. *The Journal of Marketing*, 11 (4), 390-393.
- Fuller, R. G. C., and Sheehy-Skeffington, A. (1974). Effects of group laughter on responses to humorous material, a replication and extension. *Psychological Reports*, 35, 531-534.
- Furnas, G. W., Fake, C., von Ahn, L., Schachter, J., Golder, S., Fox, K., Davis, M., Marlow, C., and Naaman, M. (2006). Why do tagging systems work? In *CHI'06 Extended Abstracts on Human Factors in Computer Systems*, ACM, New York, Ny, 36-39.

- Geerts, D. (2006). Comparing voice chat and text chat in a communication tool for interactive television. In *Proceedings of NordiCHI 2006*, ACM, New York, NY, 461-464.
- Geerts, D., Cesar, P., and Bulterman, D. (2008). The implications of program genres for the design of social television systems. In *Proceedings of uxTV 2008*, ACM, New York, NY, 71-80.
- Gibbons, J. F., Kincheloe, W. R., and Down, K. S. (1977). Tutored videotape instruction: A new use of electronics media in education. *Science*, *195*, 1139-1146.
- Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., and Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, *40* (1), 84-96.
- Gough, P. J. (2007). CNN's YouTube debate draws impressive ratings. Reuters. Retrieved October 2009 from <http://www.reuters.com/article/technologyNews/idUSN2425835220070725>
- Gough, P. J., Leffler, R., Turner, M., and Roxborough, S. (2009). Obama's inauguration most-watched since Regan's. The Live Feed. Retrieved October 2009 from <http://www.thrfeed.com/2009/01/barack-obama-inauguration-ratings.html>
- Graham, J. (2008). YouTube tosses 10-minute limit to show full TV episodes. USA Today. Retrieved October 2009 from http://www.usatoday.com/money/media/2008-10-23-youtube-tv-episodes_N.htm
- Grigonis, R. (2008). Introducing P2P-Next, a European Union Internet data and TV standard. TMCnet. Retrieved March 2008 from <http://iptv.tmcnet.com/topics/iptv-technology/articles/21510-introducing-p2p-next-european-union-internet-data-tv.htm>
- Halverson, C., Newswanger, J., Erickson, T., Wolf, T., Kellogg, W. A., Laff, M., and Malkin, P. (2001). WorldJam: Supporting talk among 50,000+. Poster at the *European Conference on Computer-Supported Cooperative Work (ECSCW 2001)*.

- Hampel, R. and Baber, E. (2003). Using internet-based audio-graphic and video conferencing for language teaching and learning. In U. Felix (Ed.), *Language Learning Online: Towards Best Practice*. Lisse: Swets & Zeitlinger.
- Handel, M. and Herbsleb, J. D. (2002). What is chat doing in the workplace? In *Proceedings of CSCW 2002*, ACM, New York, NY, 1-10.
- Harboe, G., Massey, N., Metcalf, C., Wheatley, D., and Romano, G. (2008a). The uses of social television. *ACM Computers in Entertainment*, 6 (1), 1-15.
- Harboe, G., Metcalf, C. J., Bentley, F., Tullio, J., Massey, N., and Romano, G. (2008b). Ambient social TV: Drawing people into a shared experience. In *Proceedings of CHI 2008*, ACM, New York, NY, 1-10.
- Harboe, G., Metcalf, C., Huang, E., Novak, A., Massey, N., Romano, G., and Tullio, J. (2009). Getting to know social television: One team's discoveries from library to living room. In P. Cesar, D. Geerts and K. Chorianopoulos (Eds.), *Social Interactive Television: Immersive Shared Experiences and Perspectives*. Hershey, PA: IGI Global.
- Harper, F. M., Sen, S., and Frankowski, D. (2007). Supporting social recommendations with activity-balanced clustering. In *Proceedings of RecSys 2007*, ACM, New York, NY, 165-168.
- Harrison, C. and Amento, B. (2007). CollaboraTV: Using asynchronous communication to make TV social again. In A. Lugmayr & P. Golebiowski (Eds.), *Adjunct Proceedings of EuroITV 2007* (pp. 218-222). Tampere, Finland: Tampere International Center for Signal Processing.
- Heckner, M., Neubauer, T., and Wolff, C. (2008). Tree, funny, to_read, google: What are tags supposed to achieve? a comparative analysis of user keywords for different digital resource types. In *Proceedings of the 2008 ACM Workshop on Search in Social Media*, ACM, New York, NY, 3-10.
- Henderson, R., Mahar, D., Saliba, A., Deane, F., and Napier, R. (1998). Electronic monitoring systems: An examination of physiological activity and task performance within a simulated keystroke security and electronic performance monitoring system. *Int. J. Human-Computer Studies*, 48, 143-157.

- Hills, P. and Argyle, M. (1998). Positive moods derived from leisure and their relationship to happiness and personality. *Personality and Individual Differences*, 25 (3), 523-535.
- Horrigan, J. B. and Smith, A. (2007). Home broadband adoption 2007. Pew Internet and American Life Project. Retrieved March 2008 from http://www.pewinternet.org/PPF/r/217/report_display.asp
- Huang, C. and Liao, B. (2001). A robust scene-change detection method for video segmentation. *IEEE Transactions on Circuits and Systems for Video Technology*, 11 (12), 1281-1288.
- Isaacs, E., Morris, T., and Rodriguez, T. K. (1994). A forum for supporting interactive presentations to distributed audiences. In *Proceedings of CSCW 1994*, ACM, New York, NY, 405-416.
- Isaacs, E., Morris, T., Rodriguez, T. K., and Tang, J. C. (1995). A comparison of face-to-face and distributed presentations. In *Proceedings of CHI 1995*, ACM, New York, NY, 354-361.
- Isaacs, E., Kamm, C., Schiano, D. J., Walendowski, A., and Whittaker, S. (2002). Characterizing instant messaging from recorded logs. In *CHI '02 Extended Abstracts on Human Factors in Computing Systems*, ACM, New York, NY, 720-721.
- Jancke, G., Grudin, J., and Gupta, A. (2000). Presenting to local and remote audiences: Design and use of the TELEP system. In *Proceedings of CHI 2000*, ACM, New York, NY, 384-391.
- Jensen, C., Farnham, S. D., Drucker, S. M., and Kollock, P. (2000). The effect of communication modality on cooperation in online environments. In *Proceedings of CHI 2000*, ACM, New York, NY, 470-477.
- Joinson, A. (2008). 'Looking at', 'looking up' or 'keeping up with' people? Motives and uses of Facebook. In *Proceedings of SIGCHI 2008*, ACM, New York, NY, 1027-1036.

- Jones, Q., Moldovan, M., Raban, D., and Butler, B. (2008). Empirical evidence of information overload constraining chat channel community interactions. In *Proceedings of CSCW 2008*, ACM, New York, NY, 323-332.
- Joyce, R., and Gupta, G. (1990). Identity authentication based on keystroke latencies. *Communications of the ACM*, 33 (2), 168-176.
- Karau, S. J., and Williams, K. D. (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality & Social Psychology*, 65 (4), 681-706.
- Kauff, P. and Schreer, O. (2002). An immersive 3D video-conferencing system using shared virtual team user environments. In *Proceedings of CVE 2002*, ACM, New York, NY, 105-112.
- Keele, S. W. (1973). Attention and human performance. Pacific Palisades, CA: Goodyear.
- Kellogg, W. A., Erickson, T., Wolf, T., Levy, S., Christensen, J., Sussman, J., and Bennett, W. E. (2006). Leveraging digital backchannels to enhance user experience in electronically mediated communication. In *Proceedings of CSCW 2006*, ACM, New York, NY, 451-454.
- Kiesler, S., Siegel, J., and McGuire, T. W. (1984). Social psychological aspects of computer-mediated communication. *American Psychologist*, 39 (10), 1123-1134.
- Kiesler, S. and Sproull, L. (1992). Group decision making and communication technology. *Organizational Behavior and Human Decision Processes*, 52, 96-123.
- Kittur, A., Chi, E. H., Suh, B. (2008). Crowdsourcing user studies with mechanical turk. In *Proceedings of SIGCHI 2008*, ACM, New York, NY, 453-456.
- Kramer, A. D. I., Fussell, S. R., and Setlock, L. D. (2004). Text analysis as a tool for analyzing conversation in online support groups. In *Proceedings of CHI 2004*, ACM, New York, NY, 1485-1488.
- Kraut, R. E. (1995). The HomeNet Project. <http://homenet.hcii.cs.cmu.edu/>

- Kraut, R. E., Kiesler, S., Boneva, B., and Shklovski, I. (2004). Examining the impact of Internet use on TV viewing: Details make a difference. In R. Kraut, M. Brynin, and S. Kiesler (Eds). *Domesticating Information Technology*. Oxford University Press.
- Lafferty, J., McCallum, A., and Pereira, F. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the Eighteenth International Conference on Machine Learning* (pp. 282-289). San Francisco: Morgan Kaufmann.
- Lampe, C. and Johnston E. (2005). Follow the (slash) dot: effects of feedback on new members in an online community. In *Proceedings of SIGGROUP 2005*, ACM, New York, NY, 11-20.
- Lampe, C., Ellison, N., and Steinfield, C. (2006). A face(book) in the crowd: Social searching vs. social browsing. In *Proceedings of CSCW 2006*, ACM, New York, NY, 167-170.
- Law, E. and von Ahn, L. (2009a). Input-agreement: A new mechanism for collecting data using human computation games. In *Proceedings of CHI 2009*, ACM, New York, NY, 1197-1206.
- Law, E., von Ahn, L., and Mitchell, T. (2009b). Search war: A game for improving web search. In *Proceedings of the ACM SIGKDD Workshop on Human Computation*, ACM, New York, NY, 31-31.
- Lawson, T. J., Downing, B., and Cetola, H. (1998). An attributional explanation for the effect of audience laughter on perceived funniness. *Basic and Applied Social Psychology*, 20, 243-249.
- Lee, B. and Lee, R. S. (1995). How and why people watch TV: Implications for the future of interactive television. *Journal of Advertising Research*, 35 (6), 9-18.
- Lee, M., Dillahunt, T., Pendleton, B., Kraut, R., and Kiesler, S. (2009). Tailoring websites to increase contributions to online communities. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems*, ACM, New York, NY, 4003-4008.

- Leonard, J., Riley, E., and Staman, E. M. (2003). Classroom and support innovation using IP video and data collaboration techniques. In *Proceedings of the 4th Conference on Information Technology Curriculum (CITC4)*, ACM, New York, NY, 142-150.
- Liu, Y., Shafton, P., Shamma, D. A., and Yang, J. (2007). Zync: The design of synchronized video sharing. In *Proceedings of DUX 2007*, ACM, New York, NY, 1-8.
- Löber, A., Schwabe, G., and Grimm, S. (2007). Audio vs. chat: The effects of group size on media choice. In *Proceedings of the 40th Hawaii International Conference on System Sciences*, IEEE, 1-10.
- Lochner, K., Kawachi, I., and Kennedy, B. P. (1999). Social capital: a guide to its measurement. *Health & Place*, 5, 259-270.
- Ludford, P. J., Cosley, D., Frankowski, D., and Terveen, L. (2004). Think different: Increasing online community participation using uniqueness and group dissimilarity. In *Proceedings of CHI 2004*, ACM, New York, NY, 631-638.
- Luyten, K., Thys, K., Huypens, S., and Coninx, K. (2006). Telebuddies: Social stitching with interactive television. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems*, ACM, New York, NY, 1049-1054.
- Macgregor, G. and McCulloch, E. (2006). Collaborative tagging as a knowledge organisation and resource discovery tool. *Library Review*, 55 (5), 291-300.
- Madden, M. (2007). Online video. Pew Internet and American Life Project. Retrieved March 2008 from http://www.pewinternet.org/PPF/r/219/report_display.asp
- McCleary, R. (1950). The nature of the galvanic skin response. *Psychological Bulletin*, 47, 97-117.
- McDonald, D. M. and Chen, H. (2006). Summary in context: Searching versus browsing. *ACM Transactions on Information Systems*, 24 (1), 111-141.
- McKenna, K. Y. A., Green, A. S., and Gleason, M. E. J. (2002). Relationship formation on the Internet: What's the big attraction? *Journal of Social Issues*, 58 (1), 9-31.

- Melber, A. (2008). Obama's YouTube speech tops TV ratings. *The Nation*. Retrieved October 2009 from http://www.thenation.com/blogs/state_of_change/302543
- Milgram, S. (1977). *The individual in a social world: Essays and experiments*. Reading, Mass: Addison-Wesley Pub. Co.
- Millard, W. J. (1992). A history of handsets for direct measurement of audience response. *Int. Journal of Public Opinion Research*, 4 (1), 1-17.
- Miller, B. N., Albert, I., Lam, S. K., Konstan, J. A., and Riedl, J. (2003). MovieLens unplugged: Experiences with an occasionally connected recommender system. In *Proceedings of IUI 2003*, ACM, New York, NY, 263-266.
- Miyamori, H., Nakamura, S., and Tanaka, K. (2005). Generation of views of TV content using TV viewers' perspectives expressed in live chats on the web. In *Proceedings of MULTIMEDIA 2005*, ACM, New York, NY, 853-861.
- Moray, N. (1969). *Listening and attention*. Baltimore: Penguin.
- Mu, X., Marchionini, G., and Pattee, A. (2003). The interactive shared educational environment: User interface, system architecture and field study. In *Proceedings of the 3rd ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL)*, Houston, TX, 291-300.
- Nardi, B. A., Whittaker, S., and Bradner, E. (2000). Interaction and outeraction: Instant messaging in action. In *Proceedings of CSCW 2000*, ACM, New York, NY, 79-88.
- Nathan, M., Harrison, C., Yarosh, S., Terveen, L., Stead, L., and Amento, B. (2008). CollaboraTV: Making television viewing social again. In *Proceedings of uxTV 2008*, ACM, New York, NY, 85-94.
- Nenkova, A., Vanderwende, L., and McKeown, K. (2006). A compositional context sensitive multi-document summarizer: Exploring the factors that influence summarization. In *Proceedings of SIGIR 2006*, ACM, New York, NY, 573-580.
- Nie, N. H. and Hillygus, D. S. (2002). The impact of Internet use on sociability: Time-diary findings. *IT & Society*, 1 (1), 1-20.

- Nonnecke, B. and Preece, J. (2000). Lurker demographics: Counting the silent. In *Proceedings of CHI 2000*, ACM, New York, NY, 73-80.
- Norman, D. (1968). Toward a theory of memory and attention. *Psychological Review*, 75, 522-536.
- Norris, P. (1996). Does television erode social capital? A reply to Putnam. *PS: Political Science and Politics*, 29 (3), 474-480.
- Oldenburg, R. (1999). *The great good place: Cafés, coffee shops, bookstores, bars, hair salons, and other hangouts at the heart of a community*. New York: Marlowe.
- Olguin, C. and Kruper, J. (2004). Ellis Auditorium: The design of a scalable, fun and beautiful, socializing webcast experience. Interactive Poster at Conference on Computer Supported Cooperative Work. Chicago, Illinois: ACM, November 6-10, 2004.
- Oliver, P. E. and Marwell, G. (1988). The paradox of group size in collective action: A theory of the critical mass. II. *American Sociological Review*, 53 (1), 1-8.d
- O'Neill, J. and Martin, D. (2003). Text chat in action. In *Proceedings of SIGGROUP 2003*, ACM, New York, NY, 40-49.
- Pai, V., Kumar, K., Tamilmani, K., Sambamurthy, V., and Mohr, A. E. (2005). Chainsaw: Eliminating trees from overlay multicast. In *Proceedings of IPTPS*, Ithaca, New York.
- Parker, E. A., Lichtenstein, R. L., Schultz, A. J., Israel, B. A., Schork, M. A., Steinman, K. J., and James, S. A. (2001). Disentangling measures of individual perceptions of community social dynamics: Results of a community survey. *Health Education & Behavior*, 28 (4), 462-486.
- Parkes, A. M. and Coleman, N. (1990). Route guidance systems: A comparison of methods of presenting directional information to the driver. In E. J. Lovesey (Ed.), *Contemporary Ergonomics*. London: Taylor & Francis, 480-485.

- Parks, M. R. and Roberts, L. D. (1998). 'Making moosic': The development of personal relationships on line and a comparison to their off-line counterparts. *Journal of Social and Personal Relationships*, 15 (4), 517-537.
- Pennebaker, J. W., Francis, M. E., and Booth, R. J. (2001). Linguistic inquiry and word count: LIWC (2nd ed.) [Computer software]. Mahwah, NJ: Erlbaum.
- Pesämaa, L., Ebeling, H., Kuusimäki, M. L., Winblad, I., Isohanni, M., and Moilanen, I. (2007). Videoconferencing in child and adolescent psychiatry in Finland - an inadequately exploited resource. *Journal of Telemedicine and Telecare*, 13, 125-129.
- Platow, M. J., Haslam, S. A., Both, A., Chew, I., Cuddon, M., Goharpey, N., Maurer, J., Rosini, S., Tsekouras, A., and Grace, D. M. (2005). "It's not funny if they're laughing": Self-categorization, social influence, and responses to canned laughter. *Journal of Experimental Social Psychology*, 41, 542-550.
- Porter, M. F. (1980). An algorithm for suffix stripping. *Program*, 14 (3), 130-137.
- Powazek, D. M. (2002). Design for community: The art of connecting real people in virtual places. Indianapolis, IN: New Riders Publishing.
- Preece, J. and Maloney-Krichmar, D. (2003). Online communities: Focusing on sociability and usability. In J. Jacko and A. Sears (Eds.), *Handbook of Human-Computer Interaction*. Mahwah, NJ: Erlbaum.
- Prentice, D. A., Miller, D. T., Lightdale, J. R. (1994). Asymmetries in attachments to groups and to their members: Distinguishing between common-identity and common-bond groups. *Personality and Social Psychology Bulletin*, 20 (5), 484-493.
- Price, J. L. and Mueller, C. W. (1986). *Handbook of organizational measurement*. Marshfield, MA: Pitman.
- Putnam, R. D. (1993). *Making democracy work: Civic traditions in modern Italy*. Princeton, NJ: Princeton University Press.
- Putnam, R. D. (1995). Tuning in, tuning out: The strange disappearance of social capital in America. *PS: Political Science and Politics*, 28 (4), 664-683.

- Putnam, R. D. (2001). *Bowling alone: The collapse and revival of American community*. Simon & Schuster.
- Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. In *Proceedings of IEEE*, 77 (2), 257-285.
- Radloff, L. S. (1977). The CES-D scale: A self report depression scale for research in the general population. *Applied Psychological Measurement*, 1, 385-401.
- Raghunathan, R. and Corfman, K. (2006). Is happiness shared doubled and sadness shared halved? social influence on enjoyment of hedonic experiences. *Journal of Marketing Research*, 43, 386-394.
- Ramanathan, S. and McGill, A. L. (2007). Consuming with others: Social influences on moment-to-moment and retrospective evaluations of an experience. *Journal of Consumer Research*, 34, 506-524.
- Regan, T. and Todd, I. (2004). Media center buddies: Instant messaging around a media center. In *Proceedings of NordiCHI 2004*, ACM, New York, NY, 141-144.
- Ren, Y., Kraut, R., and Kiesler, S. (2007). Applying common identity and bond theory to design of online communities. *Organization Studies*, 28 (3), 377-408.
- Resnick, P. (2002). Beyond bowling together: Sociotechnical capital. In J. M. Carroll (Ed.), *HCI in the New Millennium*, Addison-Wesley.
- Sacks, H., Schegloff, E. A., and Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50 (4), 696-735.
- San Pedro, J., Kalnikaite, V., and Whittaker, S. (2009). You can play that again: Exploring social redundancy to derive highlight regions in videos. In *Proceedings of IUI 2009*, ACM, New York, NY, 469-473.
- Sassenberg, K. (2002). Common bond and common identity groups on the Internet: Attachment and normative behavior in on-topic and off-topic chats. *Group Dynamics: Theory, Research and Practice*, 6 (1), 27-37.
- Scholl, J., McCarthy, J., and Harr, R. (2006). A comparison of chat and audio in media rich environments. In *Proceedings of CSCW 2006*, ACM, New York, NY, 323-332.

- Schulzrinne, H., Casner, S., Frederick, R., and Jacobson, V. (1996). Rtp: A transport protocol for real-time applications. RFC 1886.
- Sen, S., Lam, S. K., Rashid, A., Cosley, D., Frankowski, D., Osterhouse, J., Harper, F. M., and Riedl, J. (2006). tagging, communities, vocabulary, evolution. In *Proceedings of CSCW 2006*, ACM, New York, NY, 181-190.
- Sen, S., Harper, F. M., LaPitz, A., and Riedl, J. (2007). The quest for quality tags. In *Proceedings of SIGGROUP 2007*, ACM, New York, NY, 361-370.
- Shamma, D. A., Shaw, R., Shafton, P. L., and Liu, Y. (2007). Watch what I watch: Using community activity to understand content. In *Proceedings of MIR 2007*, ACM, New York, NY, 275-284.
- Shamma, D. A., Bastéa-Forte, M., Joubert, N., and Liu, Y. (2008). Enhancing online personal connections through the synchronized sharing of online video. In *CHI '08 Extended Abstracts on Human Factors in Computing Systems*, ACM, New York, NY, 2931-2936.
- Shen, D., Yang, Q., Sun, J., and Chen, Z. (2006). Thread detection in dynamic text message streams. In *Proceedings of SIGIR 2006*, ACM, New York, NY, 35-42.
- Shirky, C. (2005). Ontology is overrated. Retrieved September 2009 from http://shirky.com/writings/ontology_overrated.html
- Shklovski, I., Kraut, R., and Rainie, L. (2004). The Internet and social participation: Contrasting cross-sectional and longitudinal analyses. *Journal of Computer-Mediated Communication*, 10 (1). Retrieved March 2008 from http://jcmc.indiana.edu/vol10/issue1/shklovski_kraut.html
- Siegel, J., Dubrovsky, V., Kiesler, S., and McGuire, T. W. (1986). Group processes in computer-mediated communication. *Organizational Behavior and Human Decision Processes*, 37 (2), 157-187.
- Siegler, M. (2009). Facebook launches a live stream box, partners with Ustream. Techcrunch. Retrieved December 2009 from <http://www.techcrunch.com/2009/06/24/facebook-launches-a-live-stream-box-partners-with-ustream/>

- Siersdorfer, S., San Pedro, J., and Sanderson, M. (2009). Automatic video tagging using content redundancy. In *Proceedings of SIGIR 2009*, ACM, New York, NY, 395-402.
- Slater, M., Sadagic, A., and Schroeder, R. (2000). Small-group behavior in a virtual and real environment: A comparative study. *Presence, Teleoperators and Virtual Environments*, 9 (1), 37-51.
- Smith, M., Cadiz, J. J., and Burkhalter, B. (2000). Conversation trees and threaded chats. In *Proceedings of CSCW 2000*, ACM, New York, NY, 97-105.
- Smith, R. B., Sipusic, M. J., and Pannoni, R. L. (1999). Experiments comparing face-to-face with virtual collaborative learning. In *Proceedings of CSCL 1999*. C. M. Hoadley and J. Roschelle (Eds.). International Society of the Learning Sciences.
- Smyth, M. M. and Fuller, R. G. C. (1972). Effects of group laughter on responses to humorous material. *Psychological Reports*, 30, 132-134.
- Spiegel, D. S. (2001). Coterie: A visualization of the conversational dynamics within IRC. Master's Thesis. Massachusetts Institute of Technology. Cambridge, MA.
- Steinkuehler, C. A. and Williams, D. (2006). Where everybody knows your (screen) name: Online games as "third places". *Journal of Computer-Mediated Communication*, 11, 885-909.
- Stone, H. R. (1990). Economic development and technology transfer: Implications for video-based distance education. In M. G. Moore (Ed.), *Contemporary Issues in American Distance Education* (pp. 231-242). Oxford, England: Pergamon.
- Stylos, J., Myers, B. A., and Yang, Z. (2009). Jadeite: Improving API documentation using usage information. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems*, ACM, New York, NY, 4429-4434.
- Suh, B., Chi, E. H., Kittur, A., and Pendleton, B. A. (2008). Lifting the veil: Improving accountability and transparency in Wikipedia with WikiDashboard. In *Proceedings of SIGCHI 2008*, ACM, New York, NY, 1037-1040.

- Sutter, J. D. (2009). Online inauguration videos set records. CNN. Retrieved October 2009 from <http://www.cnn.com/2009/TECH/01/21/inauguration.online.video/index.html>
- Utiyama, M. and Isahara, H. (2001). A statistical model for domain-independent text segmentation. In *Proceedings of the 39th Annual Meeting on Association For Computational Linguistics*, Association for Computational Linguistics, 499-506.
- Viégas, F. B. and Donath, J. S. (1999). Chat circles. In *Proceedings of CHI 1999*, ACM, New York, NY, 9-16.
- Vig, J., Sen, S., and Riedl, J. (2009). Tagsplanations: Explaining recommendations using tags. In *Proceedings of IUI 2009*, ACM, New York, NY, 47-56.
- Voida, A., Grinter, R. E., Ducheneaut, N., Edwards, W. K., and Newman, M. W. (2005). Listening in: Practices surrounding iTunes music sharing. In *Proceedings of SIGCHI 2005*, ACM, New York, NY, 191-200.
- von Ahn, L., Blum, M., and Langford, J. (2004a). Telling humans and computers apart automatically. *Communications of the ACM*, 47 (2), 56-60.
- von Ahn, L. and Dabbish, L. (2004b). Labeling images with a computer game. In *Proceedings of CHI 2004*, ACM, New York, NY, 319-326.
- von Ahn, L., Ginosar, S., Kedia, M., Liu, R., and Blum, M. (2006a). Improving accessibility of the web with a computer game. In *Proceedings of CHI 2006*, ACM, New York, NY, 79-82.
- von Ahn, L., Kedia, M., and Blum, M. (2006b). Verbosity: A game for collecting common-sense facts. In *Proceedings of SIGCHI 2006*, ACM, New York, NY, 75-78.
- von Ahn, L., Liu, R., and Blum, M. (2006c). Peekaboom: A game for locating objects in images. In *Proceedings of SIGCHI 2006*, ACM, New York, NY, 55-64.
- Vosgerau, J., Wertenbroch, K., and Carmon, Z. (2006). Indeterminacy and live television. *Journal of Consumer Research*, 32, 487-495.

- Vronay, D., Smith, M., and Drucker, S. (1999). Alternative interfaces for chat. In *Proceedings of UIST 1999*, ACM, New York, NY, 19-26.
- Wactlar, H. (2000). Informedia – search and summarization in the video medium. In *Proceedings of Imagina 2000*, Monaco, January 31 - February 2, 2000.
- Walther, J. B. (1996). Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. *Communication Research*, 23, 3-43.
- Walther, J. B. and Parks, M. R. (2002). Cues filtered out, cues filtered in: Computer-mediated communication and relationships. In M. L. Knapp & J. Daly (Eds.), *Handbook of Interpersonal Communication* (3rd ed.), Sage Publications, Inc.
- Walter, J. B. (2007). Selective self-presentation in computer-mediated communication: Hyperpersonal dimensions of technology, language, and cognition. *Computers in Human Behavior*, 23, 2538-2557.
- Weisz, J. D., Erickson, T., and Kellogg, W. A. (2006). Synchronous broadcast messaging: The use of ICT. In *Proceedings of SIGCHI 2006*, ACM, New York, NY, 1293-1302.
- Weisz, J. D., Kiesler, S., Zhang, H., Ren, Y., Kraut, R. E., and Konstan, J. A. (2007). Watching together: Integrating text chat with video. In *Proceedings of SIGCHI 2007*, ACM, New York, NY, 877-886.
- Weisz, J. D. and Kiesler, S. (2008). How text and audio chat change the online video experience. In *Proceedings of uxTV 2008*, ACM, New York, NY, 9-18.
- Weisz, J. D. (2009). Online Video as a Social Activity. In P. Cesar, D. Geerts, and K. Chorianopoulos (Eds.), *Social Interactive Television: Immersive Shared Experiences and Perspectives*. Hershey, PA: IGI Global.
- White, S. A., Gupta, A., Grudin, J., Chelsey, H., Kimberly, G., Sanocki, E. (2000). Evolving use of a system for education at a distance. In *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences*, IEEE.

- Whitney, D. (2009). 'Lost', 'SNL', 'Grey's' tops in online viewing, Nielsen says. TV Week. Retrieved October 2009 from http://www.tvweek.com/news/2009/02/lost_snl_greys_tops_in_online.php
- Wickens, C. D., Sandry, D., and Vidulich, M. (1983). Compatibility and resource competition between modalities of input, output, and central processing. *Human Factors*, 25, 227-248.
- Wickens, C. D. and Hollands, J. G. (2000). *Engineering Psychology and Human Performance* (3rd ed.). New Jersey: Prentice Hall.
- Wickens, C. D. (2002). Multiple resources and performance prediction. *Theor. Issues in Ergon. Sci*, 3 (2), 159-177.
- Williams, D., Caplan, S., and Xiong, L. (2007). Can you hear me now? The impact of voice in an online gaming community. *Human Communication Research*, 33 (4), 427-449.
- Witten, I. H. and Frank, E. (2005). *Data mining: Practical machine learning tools and techniques*, 2nd Edition. Morgan Kaufmann, San Francisco.
- Zellars, K. L., Meurs, J. A., Perrewé, P. L., Kacmar, C. J., and Rossi, A. M. (2009). Reacting to and recovering from a stressful situation: The negative affectivity-physiological arousal relationship. *Journal of Occupational Health Psychology*, 14 (1), 11-22.