SeeSaw: I See You Saw My Video Message

Gina Venolia, John C. Tang, and Kori Inkpen
Microsoft Research
Redmond WA USA
{ginav, johntang, kori}@microsoft.com

ABSTRACT
We developed a prototype called SeeSaw that explored using *reaction video* and *auto reply* to create an engaging video messaging experience. When viewing a video message, reaction video captures a video of the viewer's reaction to share back with the message sender. After finishing viewing the video message, auto reply immediately begins recording the viewer's response. A pilot study found that SeeSaw evoked conversational and authentic interactions, even though the messages were captured remotely and asynchronously. A follow-up comparative lab study found that reaction video encouraged a more conversational exchange, while both features together enhanced the authenticity of the experience. Although participants preferred the reaction video only condition, they perceived that the reaction video plus auto reply condition combined the conversationality of a video call with the flexibility of asynchronous messaging.

Author Keywords
Mobile messaging; video messaging; asynchronous communication.

ACM Classification Keywords
H.5.3 [Group and Organization Interfaces]: Asynchronous interaction

General Terms
Human Factors; Design.

CREATING ENGAGING ASYNCHRONOUS CHATS
Recent trends in communication applications have led to a burst of new video messaging tools, each with a different unique feature. For example, Vine (http://vine.co/) allows video messages that last only six seconds, and Snapchat (http://www.snapchat.com/) allows up to ten second video messages that can only be viewed for a limited time before disappearing. Such apps have become popular for sharing fun and personal messages in a rich way. Pervasive internet connectivity and widespread use of mobile devices enables viewing and sending these video messages almost anywhere. The flexibility of asynchronous messaging means people can create and view messages when it is convenient for them.

However, when people use video messages to, for example, share a funny moment that they captured, they miss out on the recipient's reaction that they would have shared if they saw it together. Users typically compensate by conveying or re-enacting their reactions through text or emoticons, such as "lol", ":-)", in their explicit replies. We wanted to explore how rich media could be used to make asynchronous messaging more naturally engaging and conversational. We wanted to understand how *reaction video* (capturing a user's reactions as they view a video message) and *auto reply* (automatically moving from viewing a message to recording a reply) could be used to create an engaging, asynchronous conversation. We built a prototype, called SeeSaw that embodied both of these features and conducted lab studies to explore the effects of reaction video and auto reply.

After reviewing related research on asynchronous messaging and reaction videos, we describe the SeeSaw prototype and discuss a pilot study that documented participants' responses to it. We then describe a lab study that examined the separate and combined effects of reaction video and auto reply. We conclude with a discussion of how these features contribute to engaging and conversational asynchronous messaging.

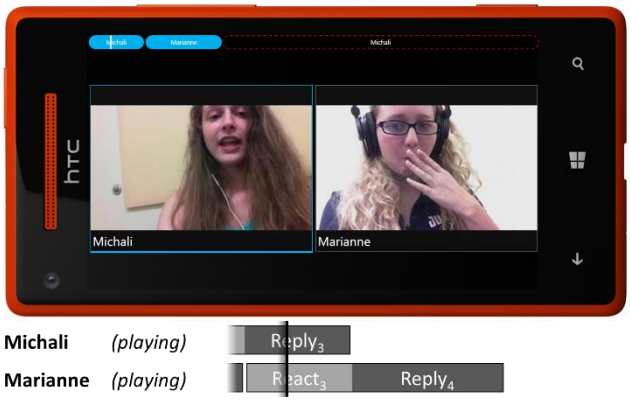

Figure 1: During the review phase, Michali views her previous outgoing video message (left) and her friend Marianne's reaction video (right).

RELATED RESEARCH
Our design approach for SeeSaw was informed by recent research on communication tools that focused on asynchronous video messaging and the use of reaction video in particular.

MobileHCI '15, August 24 - 27, 2015, Copenhagen, Denmark
 ACM 978-1-4503-3652-9/15/08…$15.00
DOI: http://dx.doi.org/10.1145/2785830.2785847

Asynchronous Video Messaging

Research on SMS texting [6] found that users appreciated the flexibility of easily sending, viewing, and responding to messages anytime, anywhere. Asynchronous texting is especially convenient when users are shifted across time zones or mobile and may not be able to attend to them immediately. Research has also documented the use of richer media, such a mobile video calls [11] and recent mobile apps that asynchronously share text, photos, and videos [12]. A recent Pew study [3] documented the rapid growth in rich media usage, noting that 54% of adult internet users have posted original video or photo content online, increasing to 81% when focusing on 18-29 year olds. Juhlin et al. [9] discussed the recent growth of video usage, fueled by the mobile device, and identified future research directions in more interactive engagement with video.

Du et al. [2] used a basic threaded video messaging system to introduce students in different countries to each other. The students enjoyed using video messages instead of email because it was richer, more expressive, and provided a stronger sense of interacting together. Inkpen et al. [7] deployed the same system in a pre-teen girl circle of friends, who used it to enthusiastically share emotions and create shared experiences together. The Family Window [8] connected homes via always-on video that included a time shift mode where families could capture a recording of their remote view if activity occurred when they were not watching. Users liked being able to view recordings of activity that they missed in real-time. Family Circles [14] explored how sharing lightweight audio messages among a family was more expressive and personal than written notes. These research projects demonstrate the potential for creating engaging sharing experiences through asynchronous video messaging.

Reaction Video

More recent projects have focused on using reaction video as a way of increasing engagement between correspondents. The Social Camera [1] used the mobile's phone front camera to capture video reactions to photos that had been shared with close friends. Users enjoyed richly sharing their emotions and phatic responses to photo sharing more than the social media "liking" gesture. KIZUNA [10] enabled dining together with globally time-shifted remote people through video recordings. A video recording was made of a remote person while he ate a meal, which was played while his friend ate her meal at a different time. Her reactions while eating were also recorded, which were played the next time he ate a meal, setting up a chain of viewings of recorded reactions around the specific activity of eating a meal. Play with Elmo [13] used the animated Elmo character to prompt video recordings by adults that elicited playful reactions from kids when they watched them. The kids' reactions were recorded and sent back to the adults so they could see if the kids sang along or copied the funny face as prompted.

In the workplace, Time Travel Proxy [15] explored supporting group meetings across the time zone differences that occur among globally distributed teams. Remote meeting participants sent their meeting contributions via a video message, and the prototype captured video reactions while meeting participants watched the video message, and continued recording to capture their responses after viewing the message. They found some rich reactions and non-verbal communication captured in video, especially when the sender explicitly included a joke. Moreover, by continuing to record the group after viewing the message, their responses were naturally recorded to give feedback to the sender.

These prior projects demonstrated the expressiveness and flexibility afforded by asynchronous messaging within specific contexts. In addition to reaction video, these projects relied on other resources in the setting, such as food in dining, or the children's book in playing with kids as part of the experience. We designed SeeSaw to explore reaction videos for a more general context on a mobile device.

THE SEESAW PROTOTYPE

We developed the SeeSaw prototype for exchanging video messages that includes both reaction video and auto reply, as shown in Figures 1-3. Interacting with an incoming message in SeeSaw consists of three phases:

- *Review*: You review the previous message you sent, with your partner's reaction playing side-by-side.
- *React*: You view the new video message your partner sent to you, during which your reactions are captured to share back with your partner.
- *Reply*: You record a reply, which starts automatically after the end of viewing and reacting to your partner's message and adds a new message to the thread with your partner.

The three phases happen in sequence without pausing between them. Note that the first couple messages of a thread are slightly different because there are no previous messages for review or reaction. Although we have used SeeSaw with groups of up to four people, for simplicity this paper focuses on how pairs of participants use it. We describe each of the messaging phases in more detail below in the context of exchanging messages between you and your partner.

Review

During the review phase, you watch the previous message that you sent to your partner along with their recorded reactions to that message. The reaction video helps convey how your partner received and reacted to your video message. Did they laugh at your joke? Did they understand what you were trying to say? Did they agree or disagree? Were they paying attention when viewing the message?

Your partner's reaction video is played alongside your original message for context; audio from both plays simultaneously. A segmented "progress bar", located at the top of the screen, provides a representation of the messages being played (one segment per message). Segments for

previously recorded messages are filled in blue. As the messages play, a cursor moves along the bar to show the replay progress. Once the review is completed, the interface moves immediately to the react phase.

React

During the react phase, shown in Figure 2, your reactions are captured while viewing the most recent message that your partner sent. While watching your partner's message (shown with a blue highlight), the microphone and camera on your device record your reaction. The video preview of your reactions being recorded (shown with a red highlight) is displayed alongside your partner's message, so you have a sense of how your reactions are being captured. The cursor continues to progress through your partner's message in the progress bar. Once you have completed viewing your partner's message (and capturing your reaction), the interface moves immediately to the reply phase. This immediate transition to replying reflects what happens in conversational turn-taking, where when I stop talking, you are on the spot to take the next turn.

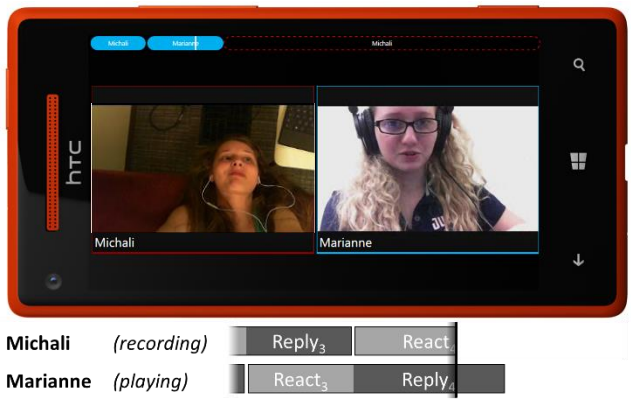

Figure 2: During the react phase, Marianne's incoming video message is shown on the right and a live video preview of Michali's reactions being recorded is shown on the left.

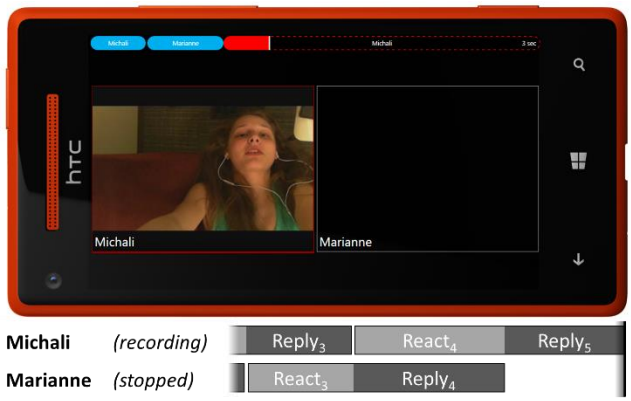

Figure 3: During the reply phase, only the live video preview of the Michali's new message being recorded is shown (left).

Reply

In the reply phase, shown in Figure 3, you record a new video to add to the thread of messages. The camera and microphone, already recording in the react phase, remain on during the reply phase. While the reply often pertains directly to the previous message from your partner, you are free to change to any topic, as in a next turn of a conversation. During this phase, you only see your video preview while you are recording; the other pane that previously showed your partner turns blank. When you have completed recording your message, tapping the progress bar at the top ends the recording and sends the message to your partner.

Implementation

Besides the interface for exchanging messages described above, there was also a screen that showed all the SeeSaw message threads that were currently active. The threads were identified by the people involved in the message exchange. Threads with a new message were marked with a star icon at the beginning. When selecting a thread with a new message, you would start reviewing at your partner's reaction to your most recent message leading up to their new message to you.

We implemented SeeSaw on a Microsoft Surface Pro tablet computer. As shown in Figure 1, the SeeSaw interface was rendered as a mobile phone, which is the target platform for this concept. Note how the pilot study participant has moved around with the device among Figures 1-3 as afforded by the prototype's mobile form factor. The prototype was implemented in C# on the .NET Framework. It shared video files and conversation state via a file server.

SEESAW PILOT STUDY

We ran a pilot study to get feedback on the combination of reaction video and auto reply features of SeeSaw. We recruited 23 participants organized into 10 sessions. We had 15 female and 8 male participants grouped into 4 female pairs, 2 male pairs, 2 mixed pairs (couples), 1 female triad, and 1 mixed group of four. All participants were 14-26 years old (median 17), an age range where current mobile messaging tools are popular. They all were active users of mobile smartphone and video apps, engaging at least weekly in video calling (e.g., Skype, FaceTime) and sharing video messages (e.g., Snapchat, YouTube) and using mobile messaging tools (e.g., SMS, WhatsApp) at least daily. In this paper, participants are identified by group number, participant number, gender and age (e.g., G1.1, F14).

We wanted to observe communication among people who knew each other. To do this we recruited one person who in turn recruited study partners with whom they actively exchanged messages. Thus, we were getting feedback on SeeSaw from intact social circles who actively used current tools for messaging with each other. Recruiting was accomplished through a mix of a professional recruiting service (15) and personal social networks (8).

Pilot Study Method

The pilot study was conducted in labs configured to look like a living room (sofa, coffee table, no computers). We first collected survey and interview data about their current communication practices: what tools and devices they used,

recent examples of video calling and video messaging, how frequently they corresponded, and so on.

We then separated the participants into individual rooms and demonstrated the SeeSaw prototype by having them send messages to each other. This familiarized them with how the user interface worked, plus got them into the typical review-react-reply pattern for SeeSaw messages. They then used SeeSaw to send messages freely to each other for about 20 minutes. They were asked beforehand to bring something to visually show each other during the session (a "show and tell" item) to help prime the messaging conversations, although not all groups showed their item in their session. We enforced a round-robin sequence of exchanging messages to ensure even participation.

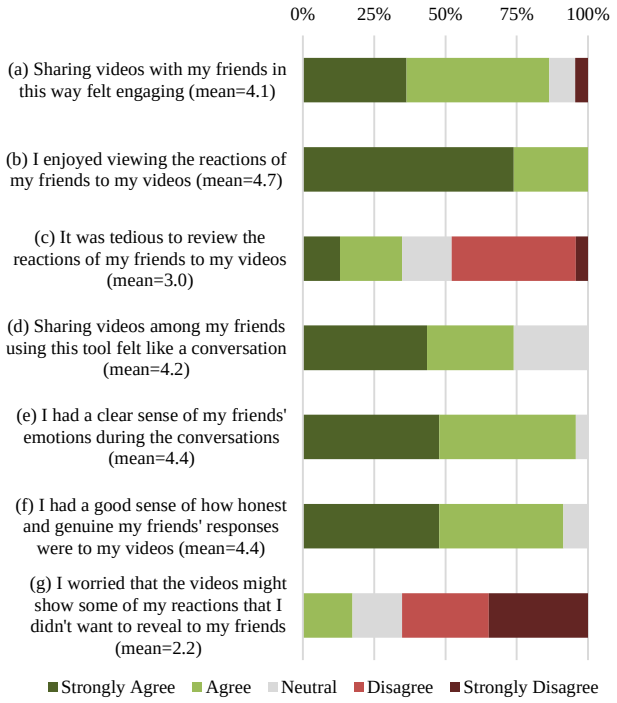

Figure 4: Responses to the pilot study post-test questionnaire Likert scale questions.

To simulate an asynchronous messaging setting within a lab study, we invited the participants to engage in other activities while they were waiting for their video responses. The rooms were stocked with snacks, popular magazines, puzzles, and toys, and they were invited to use their smartphones to pass the time while waiting for the next message. Several responses in our study were delayed by finishing a turn within a toy game or reading a paragraph before attending to SeeSaw to look at a new message. We did not constrain or suggest the topics of their conversation in any way, and many talked about artifacts in the room or ongoing social issues (what's happening at school, homework, planning, etc.)

After using SeeSaw for about 20 minutes, we administered a survey to collect their reactions to the prototype (5-point Likert scale and open-ended questions) and brought them together again for a semi-structured group interview about their experiences. The study session lasted about one hour, and each participant was given a $100 gift card (with $25 extra for the two people who recruited more than one friend).

Results

The pilot study participants had a strong positive reaction to SeeSaw, as evidenced by their engaged activity during the study, survey responses, and their comments during the interview. Participants easily filled the time through video messages with rich reactions, and 86% agreed or strongly agreed that sharing video messages through SeeSaw was engaging (Figure 4a). The nature of the interactions through SeeSaw were strikingly conversational, even though the messages were recorded and viewed asynchronously. We describe two examples that are included in the video figure.

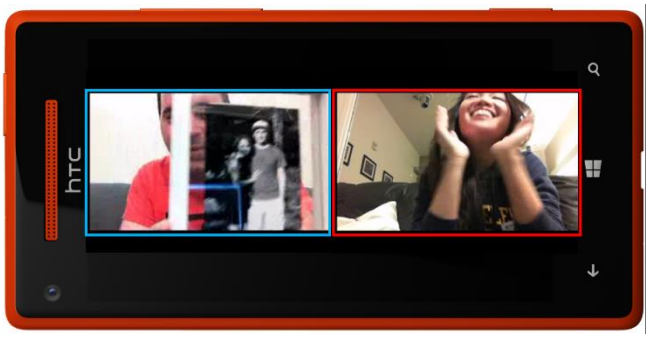

Figure 5: Husband (left) recorded a message showing a gift, which wife (right) reacts to during her viewing of his message.

One married couple (G6.1, M25 and G6.2, F22) brought surprise gifts that they revealed to each other through the SeeSaw video messages. Figure 5 shows the wife's reaction as she reviewed her husband's message showing the gift. It is a picture of when they first met, but she does not like that particular picture. The husband clearly got her reaction when he reviewed her message. SeeSaw conveyed her reaction in a way that seemed like they were having an interactive conversation, even though the messages were recorded remotely and asynchronously from each other.

Figure 6: Girl on the left recorded a message giving her friend (right) time to guess her high game score during her reaction.

In another example, shown in Figure 6, one high school girl (G1.1, F14) paused while recording a message to create a guessing game, knowing that she would hear her friend's guesses in the reaction video. She asked her friend to guess her high score on a game, filling the pause with a game show

waiting tune. Her friend hurriedly shouted out guesses until time was up (when she stopped singing the tune). This game naturally emerged after just a few turns of using SeeSaw.

These two examples illustrate the highly interactive nature of SeeSaw interactions, which seemed more like conversations despite the fact that they were asynchronous messages. Participants mentioned reaction video, conversational nature, authenticity, and asynchronous conversation as aspects of SeeSaw that they liked, as discussed below.

Reaction video

In a freeform question we asked participants to comment on their favorite feature of SeeSaw. Reaction video was the most frequently mentioned feature, with 16 of the 23 participants (70%) stating that it was their favorite feature. Additionally, all participants strongly agreed or agreed that they enjoyed viewing their friend's reactions (Fig. 4b). This interest was also reflected in the interviews:

> *I liked seeing the reactions… I can see what she was thinking when I was saying, I heard her talk back to some things while my video was playing.* G9.1, F25

> *That's what I would get it for, just to see reactions. 'Cause I like sending funny videos… the reactions, for me to see it is rewarding.* G4.1, M25

Besides the explicit reactions conveyed in the videos, it provided a gift of attention back to the sender of the message. While this was beneficial for some people, others did not like reviewing the reaction video and re-watching their previous message. 35% agreed or strongly agreed that it felt tedious to review the reaction videos (Fig. 4c). We suspect that this perception may have been related to longer messages and the artificial nature of the lab setting, given that they had just sent the message and could easily remember what they said.

Conversational nature

Participants remarked on how conversational the interactions felt, even though it was accomplished through exchanging video messages back and forth. 74% of participants agreed or strongly agreed that sharing videos in this way felt like a conversation (Fig. 4d).

> *It felt more like a conversation… like in real life. I could see his reactions to what I'm saying and stuff.* G8.1, M17

> *Felt like an in-person conversation sort of mixed with Snapchat.* G10.1, M16

The combination of reaction video, which provided a rich sense of how people were reacting to the message, and the auto reply, which kept the messages flowing, contributed to the conversational feel of exchanging messages.

Authenticity

A recurring theme emerged from participants' comments that the messages seemed authentic and genuine, similar to face-to-face conversation. 96% strongly agreed or agreed that they had a clear sense of their friends' emotion (Fig 4e) and 91% that they had a good sense of how honest and genuine their responses were (Fig. 4f).

> *It's not like a text [where] if I tell you something, I don't know how guarded your response is.* G6.2, F22

> *…you don't have time to make up something. You can't like lie about the question asked about you.* G2.4, M16

> *Because if you could stop it then… your initial reaction would be different than what you recorded. It would be more like just recording your answer if you could stop it, rather than your reaction.* G4.2, F24

The reaction video and auto reply features together appeared to contribute to this sense of authenticity. The reaction video richly conveyed reactions while viewing the message, and the immediate reply simulated face-to-face conversation where even pauses in responding could be telling.

In fact, 17% of participants registered some concern that the videos might show reactions that they did not want to reveal (Fig. 4g). These concerns evoke Goffman's distinctions of front stage and back stage behaviors [4]. Traditional messaging, especially if using less rich media such as text or photos, gives users control to compose and manage their response in the back stage before sending it. But reaction video and auto reply together create a front stage visibility similar to face-to-face conversation, even though the messages are created asynchronously and remotely.

Asynchronous conversation

The affordances of asynchronous messaging were mentioned as a favorite feature by 22% of the participants. Given that the interaction felt conversational, participants liked the flexibility and efficiency of doing it asynchronously.

> *I would probably even use this one more often than using Skype. I mean, we're pretty busy, and there's not a lot time… I can do it when it's convenient and then they can do it when it's convenient, and we can carry on a conversation that way.* G9.1, F25

> *I can easily do what I need to do in between the messages, but it's still like I'm talking to her.* G4.2, F24

We were surprised to hear participants compare SeeSaw, an asynchronous messaging tool, with real-time communication tools like Skype. While the benefits of asynchronous messaging have been largely popularized through texting and mobile messaging, SeeSaw enabled those advantages while still preserving the rich sense of having a conversation. Participants liked being able to multi-task between messages.

COMPARATIVE LAB STUDY

While we were encouraged by the conversational nature of the interactions with SeeSaw, and the strong positive feedback received from participants in the pilot study, it was unclear whether these benefits were a result of providing reaction video, or automatic reply, or the combination of both features. We wanted to better understand the benefits afforded by each of these features.

We designed a study that enabled us to examine the effects of the two features separated into four different conditions, as shown in Table 1. Condition 1 was similar to traditional video messaging tools, while condition 4 was like the SeeSaw prototype tested in the pilot study. Conditions 2 and 3 looked at auto reply and reaction video (respectively) in isolation. This design also enabled us to compare reaction video (conditions 3 and 4) with no reaction video (1 and 2) and auto reply (2 and 4) with click to reply (1 and 3). The SeeSaw prototype was reconfigured to operate in any of the four conditions of the study, so the user interface remained otherwise consistent across the four conditions.

	Click to Reply	Auto Reply
No Reaction	1. No reaction video, click to reply	2. No reaction video, auto reply
Reaction Video	3. Reaction video, click to reply	4. Reaction video, auto reply

Table 1. Four conditions studied in the comparative lab study.

Comparative Study Method

We used personal social networking to recruit 32 participants who were 14-25 years old (median 16), organized into pairs for 16 study sessions. There were 16 males and 16 females (6 female pairs, 6 male pairs, 4 mixed pairs). As before, participants actively used smartphones, mobile messaging, and video tools, and the studies were conducted in a living room setting. We again recruited one person who in turn recruited a friend to participate in the study with them. There was no overlap in participants with the pilot study.

We designed a within-subjects lab study where each pair experienced all four conditions. After the initial survey and interview on current practices, each pair tried each condition for approximately 12 turns (6 turns per person). After each condition, we administered a short questionnaire to capture their perceptions of each condition. The conditions were sequenced so that changing between them resulted in only one feature being added or subtracted. Thus the eight possible sequences were generated by starting with any condition and moving clockwise or counterclockwise around Table 1. Condition order was counter-balanced.

After completing all four conditions, a post-study questionnaire was used to gather comparison data among the four conditions, including a rank ordering of their preference among the four conditions. The pairs were then brought together for a semi-structured group interview. The entire study session lasted between 90-120 minutes, and each participant received a $125 gift card.

We made one substantive design change to the prototype after the pilot study, in addition to some usability improvements. In the pilot study, users expressed some concerns about the extra time spent reviewing each reaction video, which was especially noticeable with long video messages. The SeeSaw prototype tested in the pilot study had no time limit for the recorded video messages. Most messages during the pilot study were short (median 18.4 seconds), which was probably shaped by the limits in popular video messaging tools (10 seconds for Snapchat, 6 seconds for Vine). But several messages were longer than 60 seconds (including one that went longer than 5 minutes).

Since message length could affect how much users enjoyed the reaction video experience, we revised the interface to encourage shorter video messages. Our design provided a soft time limit that gave users feedback when their video message exceeded 20 seconds, but did not cut them off at that point. As shown in Figure 7, as users recorded their video, the white line cursor in the progress bar moved through the time interval allotted for a reply, set to 20 seconds. This interval is initially an empty frame and fills in as the cursor progressed in time. If the recording went beyond 20 seconds, the bar would be filled and the counter showing the message duration began to flash yellow and black as it increased, but it did not cut the user off. It would still be up to the user to tap the thread bar to end the recording and send the message. We hoped that the interface would encourage shorter video messages without the disruption of a hard cutoff (a common complaint in tools such as Snapchat and Instagram).

Figure 7: Shorter messages were encouraged by a 20-second long progress bar. If the recording exceeded 20 seconds, the duration counter (right side) flashed yellow and black.

Results

We present the results of the comparative lab study including three sources of data: the post-condition questionnaire, the final condition ranking, and the final interview.

The post-condition questionnaire comprised several 5-point Likert questions which asked participants to rate their level of agreement with each statement on a scale from 1=strongly disagree to 5=strongly agree. For each question, we ran a two-way repeated measures (2x2) ANOVA analysis with *reaction video* and *auto reply* as the two within-subject factors (Table 2). (Additionally we ran *gender* as a between-subject factor but found no significant results, and so we removed gender from our analysis.)

After experiencing all four conditions, participants were asked to rank each condition in order of preference (Figure 8). We performed two-way repeated measures (2x2) ANOVA analysis, with *reaction video* and *auto reply* as the two within-subject factors. (As with the Likert questions, gender had no effect and was removed from analysis.)

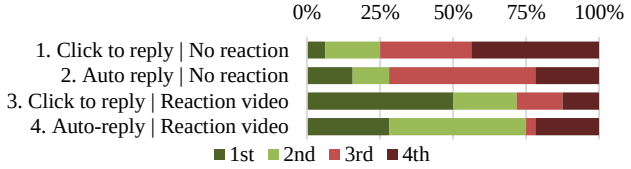

Figure 8: Participants' ranking of the four conditions.

	Auto Reply (AR)		Reaction Video (RV)		Interaction
	Significance	**Means**	**Significance**	**Means**	
(a) I had a clear sense of how my friend reacted to my videos	$F_{1,31}$=.011 p=.916	Without AR=3.9 With AR=3.9	$F_{1,31}$=51.251 ***p*<.001**	Without RV=3.3 With RV=3.5	$F_{1,31}$=.798 p=.379
(b) I could easily tell whether my friend got the humor in my videos to them	$F_{1,31}$=.252 p=.619	Without AR=4.0 With AR=4.0	$F_{1,31}$=22.776 ***p*<.001**	Without RV=3.6 With RV=4.4	$F_{1,31}$=.660 p=.423
(c) It felt more like having a conversation than sending separate messages	$F_{1,31}$=2.319 p=.138	Without AR=3.6 With AR=3.8	$F_{1,31}$=27.414 ***p*<.001**	Without RV=3.3 With RV=4.2	$F_{1,31}$=2.159 p=.152
(d) I had a good sense of how honest and genuine my friend's replies were to my videos	$F_{1,31}$=.574 p=.455	Without AR=4.0 With AR=3.9	$F_{1,31}$=30.834 ***p*<.001**	Without RV=3.6 With RV=4.4	$F_{1,31}$=.225 p=.638
(e) I worried that the videos might show some of my reactions that I didn't want to reveal to my friend	$F_{1,31}$=3.583 p=.068	Without AR=2.3 With AR=2.7	$F_{1,31}$=21.406 ***p*<.001**	Without RV=2.1 With RV=2.9	$F_{1,31}$=.377 p=.544

Table 2: ANOVA results for post-condition Likert questions (strongly disagree=1 through strongly agree=5).

All the interviews were video recorded, and analyzed using an open coding approach. We reviewed the recordings to detect recurring themes and find representative quotes.

Reaction Video

ANOVA of the overall rankings for each condition revealed a significant main effect for *reaction video* ($F_{1,30}$=10.28, p=.003) with the two reaction video conditions (3 & 4) being ranked significantly higher than the two non-reaction video conditions (1 & 2). No significant difference for *auto reply* was found ($F_{1,30}$=.03, p=.872) but there was a significant interaction effect between *reaction video* and *auto reply* ($F_{1,30}$=5.60, p=.025). Given the ordinal nature of the data, we used a Wilcoxon Signed Ranks Tests with a Bonferroni correction to explore this interaction effect. We found that without auto reply, adding reaction video significantly increased the participants' rankings (p=.001) but when auto reply was used, adding reaction video did not cause a significant increase in rankings (p=.078).

Reaction video had a significant effect on several aspects of the conversation (Table 2a-d). Interview comments explained what participants liked about the reaction video:

> *Like when people type 'lol', you're not laughing, and I know you're not laughing, so, with reactions, you know!* G14.2, F14

> *When we went from reaction to no reaction, it felt so slow and boring.* G15.2, M21

Some participants felt that reaction videos could reveal reactions that they did not want to reveal to their friend (Table 2e). 29% of participants agreed or strongly agreed that they were worried about unintentionally revealing their reactions in the reaction video conditions, compared to only 9% in the no reaction conditions.

> *I thought the reaction, while you have less control over it, it was authentic… You have less control as a viewer when you're seeing something for the first time, and you're being recorded…* G8.1, M18

A few participants commented that they did not like reviewing their own messages as part of seeing their partner's reactions, because they did not like seeing themselves or thought it was tedious.

> *I'm not really interested in seeing myself again, of how I look, I just want to forget about it.* G3.1, F16

> *In this case, his reactions weren't that interesting… kind of felt like a waste of time.* G13.2, M18

However, several participants volunteered that it would be useful to see their original messages again to restore the context about the content of the message.

> *Then I really liked the side of seeing my video again, because generally when I do Snapchat or something like that, I don't respond right away, and those are the times I do forget what I said, so I really liked that.* G17.2, F16

> *You forget, kind of exactly what you say. If you saw a reaction without your own video alongside it, I would have no context to what's going on… If I had two of these conversations going… it would be totally impossible to keep track of.* G15.2, M21

It was notable that the lab study evoked this expectation of how this feature would be useful in real life.

Auto Reply

The interviews revealed several aspects about auto reply that participants appreciated.

> *It was more like a flowing conversation for me than a messaging system. You can kind of just bounce back and forth, and there's not as much lag time between it changing screens, you having to hit the click to reply, collecting your thoughts. You just kind of went with the conversation as it would go if you were in person.* G16.1, F20

> *The auto reply is a lot more fluid, because instead of stopping at the end of the video and having to press a button so that you can respond to the person… We don't really get that with Snapchat or Vine.* G8.2, F16

> *You were more put on the spot, you had to react and reply at the same time, which was different than you usually see in those media sharing.* G8.2, F16

However, participants also commented that click to reply gave them a chance to compose themselves before sending a message. They were sometimes caught off guard with auto reply (especially since they may have recently experienced click to reply conditions) and preferred click to reply as more familiar and generating less anxiety.

> *The main thing I didn't like about the auto reply, was, while I'm watching a video, and then immediately after the conclusion, I'm put on the spot to reply.* G8.1, M18

> *I did really like how it saved time, the way that it was efficient, you didn't have to push reply or anything like that, but the little portion of time between when it starts and you start talking, and there's that little bit of a lag, I thought was a little bit awkward.* G17.2, F16

Participants also appreciated that auto reply was simpler (fewer touch gestures) and facilitated the conversation flow, but were not convinced those benefits outweighed the disadvantages. While participants preferred auto reply over traditional video messaging, it may have been overshadowed in the reaction video conditions given the participants strong preference for reaction video.

Reaction video and auto reply together

While only 28% of participants ranked the reaction video with auto reply condition as their first choice, their perceptions of this condition echoed what the pilot study participants liked about SeeSaw. A recurring theme in the interview comments was that this condition was more like a real-time video call, but with the benefits of asynchrony.

> *The auto-reply and reaction was definitely about as close to a video call as I think I've seen without being in a video call. The time commitment is less… but you're still maintaining what feels to be a continuous conversation, even though there's down time in between.* G8.1, F16

> *[The auto-reply and reaction video condition] really felt like a conversation, except that you could view it at later time. Skype's really lacking that, both of you have to be online at the same time.* G9.1, M15

> *I liked it because if I ever didn't have the time to have a call with someone, if I was doing that, then it would help me remember what I said, and how they reacted, and it could continue the conversation.* G16.1, F20

Thus, participants found the reaction video with auto reply condition to be more like an interactive conversation that could happen asynchronously. Their preference rating likely reflected the fact that some people were uncomfortable being "on the spot" in creating responses. The post-condition question about whether they felt that they inadvertently revealed unintended reactions, showed a near-significant main effect for auto reply ($p=.06$). This undesirable potential for revealing reactions can also be a concern in real-time conversation, and harkens to Goffman's distinctions of back stage and front stage behaviors [4]. While fewer participants preferred the reaction video with auto reply condition than reaction video with click to reply, they did recognize the former as being more like an interactive conversation.

Message duration

We analyzed the duration of video messages during the comparative lab study relative to the pilot study to see the impact of our soft time limit design. A two-sample t-test assuming unequal variances of log-adjusted durations found that the average message duration was significantly shorter in the lab study (14.21 sec) compared to the pilot study (29.50 sec), $t_{24}=3.38$, $p<.014$. Figure 9 shows a comparison of message duration between the pilot and lab study.

The distribution of the data show how the soft time limit encouraged shorter video messages in the lab study compared to the pilot. With no time limits during the pilot, participants self-selected to record generally short messages, but there was a tail of messages that exceeded 60 seconds. In the lab study, 81% of the messages were less than the 20 second limit, but 16% lasted between 20-30 seconds, and no messages were longer than 60 seconds. A hard time limit would have cut off 19% of messages longer than 20 seconds. We believe that the soft time limit effectively accomplished the goal of keeping messages shorter in a less disruptive way.

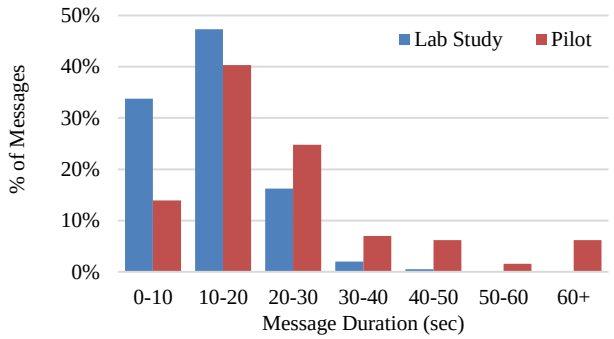

Figure 9: Comparison of message duration between the pilot and lab study.

LIMITATIONS

Studying SeeSaw in a laboratory setting using a within-subjects design adds control over unintended variables, but reduces the ecological validity of its use. In particular, the compressed timeframe of the lab study reduced the time between receiving an incoming message and viewing it. We attempted to mitigate this effect by providing distractions for the participants and encouraging them to use their mobile phones as they would normally. We observed many participants deferring viewing a message while they completed one of these activities. Furthermore, current messaging practices, especially among youth, include exchanging messages nearly synchronously for bursts of

time, similar to our lab experiment setup, so the compressed timespan of the lab study also naturally occurs in current use practices with messaging tools.

While the topical continuity of the conversations exchanged in SeeSaw may have been encouraged by the compressed timeframe of the lab study, we believe that there are important aspects of conversation interaction which were independent of the lab setting. Demonstrating attention, backchanneling, and reacting expressively were interactional mechanics [5] that we observed in our lab study as being supported by the design of SeeSaw. This conversational mechanics support is evident in the video figure and also noted in several of the user quotes cited in the Results section. We believe that SeeSaw's support for the interaction mechanics contributed to the sense of conversationality independent of the experimental setting.

DISCUSSION

Looking over the pilot study and comparative lab study together, we have more focused insights on the reaction video and auto reply features. In the pilot study, participants enjoyed using the SeeSaw prototype and found the experience to be engaging and conversational. In the comparative lab study we found that this user experience was primarily evoked by the reaction video feature. However, both the reaction video and auto reply features contributed to the authenticity benefits. Based on these results, examination of these features alone and in combination offers two designs worth exploring further.

Reaction Video Feature

We found the strongest user preference for the reaction video feature on its own. Participants found that reaction video created authentic, engaging, and fun conversations without the anxieties of a real-time conversation that seemed to be invoked by auto reply. As one participant commented on the reaction video with click to reply condition, "*I think it was like a best of both worlds combination, because you still get that conversation, but you can still take a minute to regroup and go right ahead.*" (G12.1, F15). Like prior systems that explored reaction video [1, 10, 13, 15], we found that this feature engendered engagement. In SeeSaw, reaction videos furthermore encouraged ongoing, turn-taking conversation.

Reviewing the conversational turn taking literature [5] provides some additional insight into some of the ways that reaction video could fundamentally shape the interaction. While real-time conversations are generally organized into turns of talk that do not involve much overlap, research has shown that listening during someone else's turn of talk is not a passive role. Demonstrating paying attention to the person talking can shape the conversation, including facilitating turn taking. For example, focusing eye gaze on the current speaker can prompt that person to continue talking (because the listener's attention has been established) or suggest the desire to take the next turn, depending on the context. This visual attention enables the socially constructed conversational turn taking between speakers and listeners.

While the reaction video feature was conceived as a way to give feedback to the originator about reactions to their message, we believe that it is actually doing more than that. While demonstrative reactions to video messages (e.g., laughing at a joke, expressing surprise, showing an emotional response) are memorable, they occur only occasionally during the course of typical conversation. However, the reaction video consistently gives a moment-to-moment demonstration of attention to the speaker's message. We believe that the way that reaction video affords this demonstration of attention evokes the conversational engagement that is evidenced both in the kinds of messages exchanged in the prototype and the frequent mention of feeling like a conversation in the surveys and interviews. The reaction video affords a sense of attention that generates a feel of an asynchronous conversation, rather than exchanging messages in conventional video messaging systems. While we expected that the combination of reaction video and auto reply was needed to create a conversational feel through messaging, it seems like the effect can be evoked by reaction video by itself.

Reaction Video with Auto Reply

While the reaction video with auto reply condition was not the users' most preferred condition, they did recognize that it was most like an interactive conversation that could be spread out over asynchronous sessions. We were surprised that an asynchronous messaging tool could feel like a synchronous conversation. We believe that the combination of reaction video and auto reply evokes a front stage behavior [4] and sense of turn taking (you start talking when I finish) that gives it a real-time conversational feel. This strong resemblance to interactive conversation even seems to have evoked some anxiety associated with having to reply "on-the-spot", accounting for the lower user preference.

While this condition may not be the most popular for social conversation, it may be just the right tool for situations where authentic reactions are important. For example, a tool could be created that enables asynchronous interviewing through responding to a list of questions for news stories, job applicants, police reports, medical diagnosis, or even therapeutic counseling. Enabling these interactions to occur asynchronously may provide more scheduling flexibility, especially if time zone differences are involved. Yet reaction video with auto reply could retain the real-time response dynamics, including telling pauses when answering a question, which can be crucial in many kinds of interactions. These kinds of asynchronous conversations may arise more in workplace settings and may occur among relative strangers. We might also expect that this combination of features could support sensitive or emotionally charged social communication (e.g., relationship issues, financial debates) which did not occur in the context of our lab study.

Like Time Travel Proxy [15], we found that reaction videos and auto reply combined to provide awareness for the speaker of whether their message was understood. In

SeeSaw, we found that these features furthermore created a medium that was authentic and conversational.

CONCLUSION AND FUTURE WORK

Our research identifies a couple different opportunities for creating new video messaging tools. We see design opportunities for reaction video only and in combination with auto reply. During the course of our research, several startup apps were released that also explore this space. For example, Samba (http://www.samba.me/) and Dumbstruck (http://dumbstruck.me/) capture users' reactions while viewing video messages on their mobile smartphone. Chatwala (http://chatwala.com/) captures reactions during and immediately after viewing a message.

Participants' preference for reaction video supports the recent flurry of these reaction messengers introduced in the market, especially for social communication among close ties, which was the context we studied. Beyond reactions for individual messages, which is the focus of the current startup apps, our analysis suggests that connecting those reactions together into threads will support a more conversational and interactive medium. We see an opportunity to use asynchronous messages to create a sense of synchronous conversation, blurring a line that has been used to classify CSCW research for decades. Future work could explore how SeeSaw or these apps are used in situ, exploring longitudinal effects on the relationships among their users and how they use other tools. Although we focused on studying pairs using SeeSaw, it can actually support sharing reactions among a group conversation, which would also be a promising future study topic given the recent popularity of group messaging.

Our understanding of how users perceive those configurations of features enables designers to create the tool that is the right fit for the specific situation. While our studies only looked at social communication among close ties, we see implications for future work that could explore a design space that includes social or workplace contexts involving people along a spectrum of familiarity engaged in a wide range of conversational activity.

ACKNOWLEDGMENTS

We thank all the participants of our lab studies (and the parents who brought them there).

REFERENCES

1. Yanqing Cui, Jari Kangas, Jukka Holm, and Guido Grassel, Front-camera video recordings as emotion responses to mobile photos shared within close-knit groups, *CHI 2013*, 981-990.
2. Honglu Du, Kori Inkpen Quinn, Konstantinos Chorianopoulos, Mary Czerwinski, Paul Johns, Aaron Hoff, Asta Roseway, Sarah Morlidge, John C. Tang, and Tom Gross, VideoPal: Exploring Asynchronous Video-Messaging to Enable Cross- Cultural Friendships, *ECSCW 2011*, 273-292
3. Maeve Duggan, Photo and Video Sharing Grow Online, October 28, 2013, http://www.pewinternet.org/2013/10/28/photo-and-video-sharing-grow-online/, (verified May 16, 2014).
4. Erving Goffman, *The Presentation of Self In Everyday Life*. New York: Doubleday, 1959.
5. Charles Goodwin, Turn Construction and Conversational Organization, *Rethinking Communication: Paradigm Exemplars*, Brenda Dervin, Barbara J. O'Keefe, Ellen Wartella (Eds.), Newbury Park: Sage Publications, 1989, 88-102.
6. Richard Harper, Leysia Palen, and Alex Taylor (Eds.), *The inside text: Social, cultural and design perspectives on SMS*, Dordrecht: Springer, 2005.
7. Kori Inkpen, Honglu Du, Asta Roseway, Aaron Hoff, and Paul Johns, Video kids: augmenting close friendships with asynchronous video conversations in VideoPal. *CHI 2012*, 2387-2396.
8. Tejinder K. Judge, Carman Neustaedter, and Andrew F. Kurtz, The family window: The design and evaluation of a domestic media space, *CHI 2010*, 2361-2370.
9. Oskar Juhlin, Goranka Zoric, Arvid Engström, and Erika Reponen, Video interaction: A research agenda, *Personal and Ubiquitous Computing* (18) 2014, 685-692.
10. Mamoun Nawahdah and Tomoo Inoue, Virtually dining together in time-shifted environment: KIZUNA design, *CSCW 2013*, 779-788.
11. Kenton O'Hara, Alison Black, and Matthew Lipson, Everyday practices with mobile video telephony, *CHI 2006*, 871-880.
12. Kenton O'Hara, Michael Massimi, Richard H. R. Harper, Simon Rubens, and Jessica Morris: Everyday dwelling with WhatsApp, *CSCW 2014*, 1131-1143.
13. Hayes Raffle, Rafael Ballagas, Glenda Revelle, Koichi Mori, Hiroshi Horii, Christopher Paretti, and Mirjana Spasojevic, Pop goes the cell phone: Asynchronous messaging for preschoolers, *IDC 2011*, 99-108.
14. Ruud Schatorje and Panos Markopoulos, Intra-Family Messaging with Family Circles, *Connecting Families: The Impact of New Communication Technologies on Domestic Life*, Springer, 2012, 57-74.
15. John C. Tang, Jennifer Marlow, Aaron Hoff, Asta Roseway, Kori Inkpen, Chen Zhao, and Xiang Cao, Time Travel Proxy: Using lightweight video recordings to create asynchronous, interactive meetings, *CHI 2012*, 3111-3120.