INFANT AND TODDLER ENGAGEMENT IN VIDEO MEDIATED INTERACTIONS

A dissertation
submitted to the Faculty of the
Graduate School of Arts and Sciences
of Georgetown University
in partial fulfillment of the requirements for the
degree of
Doctor of Philosophy
in Psychology

By

Elisabeth Renée McClure, M.A.

Washington, D.C.
August 14, 2015
INFANT AND TODDLER ENGAGEMENT IN VIDEO MEDIATED INTERACTIONS

Elisabeth Renée McClure, M.A.

Thesis Advisor: Yulia Chentsova-Dutton, Ph.D.

ABSTRACT

When families – especially those with very young children under 2 years old – are geographically separated, forming warm family relationships can be a challenge. Fortunately, rapid developments in communication technology are transforming our ability to interact at a distance. Families with school-aged children have reported using video chat as a method of staying in touch with remote family members; however, little is currently known about whether families with children under two are using video chat, and how often. Furthermore, few studies have evaluated whether babies and toddlers are able to use this technology successfully and whether they are emotionally engaged by it.

The present dissertation, in the form of three studies, explores the role that video chat technology plays in facilitating communication between very young children and their distant family members. In the first study, an electronic media usage survey is utilized to assess the degree to which D.C.-area families with children under 2 currently use video chat. In the second study, a naturalistic observational method is employed to examine the way families use video chat at home with their children under 2 years of age. In the third study, a controlled experimental study is used to systematically compare the emotional engagement of 6- to 12-month-olds during mother-baby interactions taking place either face-to-face, via video chat, or via non-contingent video.
The findings as a whole suggest that video chat is a promising mode of remote communication for families with babies and toddlers: it is accessible to families and used by them with regularity; when it is used, it is done fairly successfully; and the kinds of sensitive behaviors that are used with babies face-to-face can also be used to engage positively with babies via video chat.
ACKNOWLEDGEMENTS

This dissertation would not have been possible without the help and support of so many people. To you all I will always be grateful.

I would like to thank my committee, Dr. Rachel Barr, Dr. Gerrod Parrott, and Dr. Ellen Wartella, for their commitment to this work and to my development. I am humbled to have had the opportunity to learn from and work with such celebrated scholars.

Thank you to several funding sources for providing the critical resources required for these studies, including the Georgetown University Grant-in-Aid and the Psi Chi Graduate Student Research Grant.

I cannot adequately express the gratitude and warmth I feel for those on the research team who worked with me daily to collect, organize, and code data. First, thank you to Rachel Cohen, the project manager, for giving so generously of her time and support. I could not have done this without you by my side. Second, thank you to the incredible research assistants who worked tirelessly on the project: Katerina Maylock, Laura Pevytoe, Danyang Han, Amritpal Singh, Maia Sekulski, Dorosí Valle Flores, Emilia Pickard, Theresa Gladstone, Aya Kellam, Vasiliki Dimopoulos, Marianna Menakaya, Alex Grevers, Nadia Al-Taie, and Emily Fletcher. Thank you also to the lab managers at the Georgetown Early Learning Project for their recruitment support: Amanda Grenell, Herietta Lee, and Tish Tyson. Finally, thank you to Jon McClure for volunteering his time and programming expertise: you saved the day.

It has been a privilege to be surrounded by the incredible members of the Georgetown Department of Psychology over the past 5 years. I have been mentored and deeply changed by so many incredible people here. In particular, thank you again to Dr. Rachel Barr for taking me under her wing on so many occasions, providing mentorship and profound research support while asking little in return. Thank you also to Dr. Sandra Calvert for her faith and investment in my potential, for introducing me to new ideas, and for giving me a solid foundation. To Dr. Steve Sabat and Dr. James Lamiell: Your personal mentorship has changed me deeply. You are both an inspiration, and I endeavor to be more like you in my life and work.

Thank you also to the graduate students in the department for their camaraderie and support. You have made this challenge an enjoyable one, and I’m so glad to have shared this time with you. To my dear friends Calli and Erin: your friendship nourishes my heart and mind.

Finally, thank you, from the bottom of my heart, to my advisor, Dr. Yulia Chentsova-Dutton: I would not be here without your courage, compassion, and selflessness. Your gentle guidance has given me the freedom to grow and learn in the knowledge that I am supported unconditionally. Thank you for being open to this wild adventure.
This work is dedicated to my family.

You have blanketed me with love from the moment I entered the world, and I wouldn’t have been able to take these risks without the knowledge of your unreserved support. Mom, your grace and goodness inspire me daily, and your reassuring words have given me the strength to keep going. Dad, you have instilled in me your curiosity, integrity, and endurance; without these, I would have given up long ago. To my siblings: You have taught me how to learn; I am only here because I have stood on your shoulders.

Geremy, thank you for having the courage to walk this path with me. Your daily support sustains my dreams. And to baby Aden: Your sweet love reminds me I am whole.

Many thanks,

Elisabeth R. McClure
# Table of Contents

CHAPTER I: Introduction ............................................................................................................. 1

CHAPTER II: Video Chat Usage Among Infants and Toddlers

  Introduction .......................................................................................................................... 5
  Method ................................................................................................................................. 7
  Results ................................................................................................................................. 8
  Discussion ............................................................................................................................ 16
  Conclusion ........................................................................................................................... 18

CHAPTER III: Bridging the Distance: How babies and remote relatives connect using video chat at home

  Introduction ........................................................................................................................ 20
  Method ................................................................................................................................. 26
  Results ................................................................................................................................. 31
  Discussion ............................................................................................................................ 44

CHAPTER IV: Infant Emotional Engagement in Face-to-Face and Video Chat Interactions with their Mothers

  Introduction ......................................................................................................................... 54
  Method ................................................................................................................................. 59
  Design ................................................................................................................................. 62
  Results ................................................................................................................................. 68
  Discussion ............................................................................................................................ 72

CHAPTER V: General Discussion ............................................................................................. 85

Appendix A: Code Definitions and Reliability ......................................................................... 99

Appendix B: Attention Correlation Table ................................................................................ 105

References ............................................................................................................................ 106
LIST OF FIGURES

CHAPTER II
Figure 1. Most frequently used words to describe family motivations for using VMC ............ 13

CHAPTER III
Figure 2. Three types of video mediated JVA ................................................................. 28
Figure 3. Number of babies who initiated at least one instance of within-screen and/or
across-screen JVA, by age group ........................................................................ 36
Figure 4. Parents encourage the child to give his grandmother a screen kiss .................. 40
Figure 5. Baby tries to touch the device while parent pulls the device away .................. 48
Figure 6. Baby turns phone screen to show his grandmother an airplane toy (left panel),
then turns the phone to show her a block house he is working on (right panel) .......... 50
Figure 7. Sequence: Baby offering raisin through the screen ........................................ 52

CHAPTER IV
Figure 8. Baby interaction booth ................................................................................ 60
Figure 9. Mother interaction booth ............................................................................ 61
Figure 10. Unintentional facial occlusion ................................................................. 67

CHAPTER V
Figure 11. Screenshot from Kindoma’s Storytime app demonstrating the shared
pointing feature ........................................................................................................ 97
LIST OF TABLES

Table 1. Media use for children up to 24 months, over time ................................................................. 9
Table 2. Frequencies of affirmative responses for each device & probabilities of device usage by class .................................................................................................................. 10
Table 3. Average participation and role type in video mediated JVA .......................................................... 34
Table 4. Relative’s sensitivity and across-screen JVA as predictors of baby attention ......................... 42
Table 5. Correlation coefficients and significance levels for attention .................................................. 43
Table 6. Tests for non-experimental group differences ............................................................................. 68
Table 7. Tests of change score differences in physiological reactivity, by condition ......................... 70
Table 8. Latency to Hide and Condition as Predictors of Smile Frequency ........................................... 78
Chapter I: Introduction

Portions of this chapter are excerpted from a book chapter submitted as:


Modern living conditions, including labor mobility, military deployment, and parental incarceration, have led to the need for many families to maintain relationships with one another at a distance. When families – especially those with very young children under 2 years old – are geographically separated, forming warm family relationships can be a challenge. Fortunately, rapid developments in communication technology are transforming our ability to interact at a distance. Families with school-aged children have reported using video chat to help their children develop and maintain relationships with parents who are separated from them by work (Yarosh & Abowd, 2011), divorce (Yarosh, Chieh, & Abowd, 2009), or immigration (Madianou & Miller, 2012), as well as with remote grandparents (Ames, Go, Kaye, & Spasojevic, 2010).

Many families today may need to rely on methods like video chat to keep in touch. For example, as many as 10 million children in the United States experienced at least one military deployment of a loved one between 2003 and 2013; 37% have children under age 6 (American Academy of Pediatrics, 2013a; Dayton, Walsh, Muzik, Erwin, & Rosenblum, 2014). Furthermore, as of 2014, 2.7 million children in the U.S. had an incarcerated parent (National Resource Center on Children and Families of the Incarcerated, 2007); as of 2007, 92% of incarcerated parents were fathers and 22% of their children were under the age of five (National Resource Center on Children and Families of the Incarcerated, 2007). Additionally, more than 1 million children per year experience the divorce of their parents (American Academy of
Pediatrics, 2002) and 1/3 of children born in the US are born outside marriage. Of those fathers who are unmarried at the time of birth, approximately 2/3 of them no longer live with their children at age 5 and their parenting involvement varies widely, with approximately ¾ of fathers having infrequent or no contact with their children (Carlson, McLanahan, & Brooks-Gunn, 2008).

Close family relationships are known to be of critical importance for child development, especially for children under the age of three (Cassidy & Shaver, 2008), so these familial separations are important to consider. In fact, mother-infant interaction, on which most parent-infant research has been conducted, has been called “the cradle of social understanding” (Rochat & Striano, 1999), and it supports the development of attachment, communication and language acquisition, and emotion regulation (Bornstein & Tamis-LeMonda, 2001). Thus, the physical or emotional absence of a parent during the early years of a child’s life puts these infants at risk for a number of negative outcomes, including an insecure attachment to the absent parent (Tomlin, Pickholtz, Green, & Rumble, 2012).

Moreover, the number of families living in multi-generational family households in the United States has decreased since the 1940s from 25% to 16% in 2008 (Taylor, Morin, Cohn, & Wang, 2008). Circumstances like labor mobility – the change in location of workers across different jobs and geographical areas (Long & Ferrie, 2003) – can lead these extended families to live far apart from one another. According to the PEW Research Center (Taylor et al., 2008), those Americans who move away from their hometowns most often cite economic opportunity as their reason for doing so. Importantly, those who choose not to move most often cite family connections as their reason for staying (Taylor et al., 2008). Grandparents and other extended family members are considered to be very important for young children and are drawn upon
extensively in many cultures. Indeed, shifting demographics suggest that grandparents may play an increasing role in childrearing (Dunifon, 2013). Research examining the effects of non-residential grandparents on child outcomes has been mixed, but geographical distance is a significant predictor of the association between grandparents and child outcomes, with less geographical distance associated with stronger findings (Dunifon, 2013).

Can video chat be used successfully between very young children and their relatives? Because video chat offers the ability to interact contingently in real time, unlike pre-recorded videos sometimes used to assist with parenting at a distance (Yeary, Zoll, & Reschke, 2012); and because it adds a visual element to communication media like telephones, which can be difficult for young children to use (Ballagas, Kaye, Ames, Go, & Raffle, 2009), it promises to be a more effective communication medium for toddlers and their remote family members.

Few studies, however, have evaluated the engagement of babies and toddlers during video chat interactions. In fact, little is currently known about whether families with children under two are even using video chat, and how often. We do know that the prevalence of smartphone ownership among families with young children has grown in the past 5 years – 75% of families with children between 0 and 8 years of age own smartphones or some other mobile touchscreen device (Rideout, 2013) – suggesting that video chat has become more accessible. More specific documentation of video chat usage among babies and toddlers has not yet been conducted.

While relatively little is currently known about the effectiveness of video chat for supporting relationships between infants and remote family members, preliminary studies have shown that toddlers remain content for longer when they have access to a parent via video chat than when they are completely alone (Tarasuik, Galligan, & Kaufman, 2011) or when they have
access to a parent via audio-only telephone (Tarasuik, Galligan, & Kaufman, 2013). However, there is also evidence that two-dimensional media can be difficult to process for toddlers (for review see Barr, 2013), which may suggest that video chat technologies would make little difference to such young children.

Video chat appears to be a promising solution for parents and grandparents attempting to maintain connections with infants at a distance; however, with little existing research supporting or even documenting its usage, basic studies on this technology are needed. The present dissertation, in the form of three studies, will explore the role that video chat technology may play in facilitating communication between very young children and their distant family members. In the first study (Chapter II), an electronic media usage survey is utilized to assess the degree to which D.C.-area families with children under 2 currently use video chat. In the second study (Chapter III), a naturalistic observational method is employed to examine the way families use video chat at home with their children under 2 years of age. In the third study (Chapter IV), a controlled experimental study is used to systematically compare the emotional engagement of 6-to 12-month-olds during mother-baby interactions taking place either face-to-face, via video chat, or via non-contingent video.
Chapter II: Video Chat Usage Among Infants and Toddlers

A revised version of this chapter has been published as:


Introduction

Many children in the United States are geographically separated from their parents or grandparents today (American Academy of Pediatrics, 2002, 2013a, 2013b; Pew Research Center, 2010). Strong family ties are known to be critical for healthy child development (Shonkoff & Phillips, 2000), and modern communication technologies have become an important resource for families developing these bonds at a distance. Today, a military parent deployed abroad can still interact and play with his or her infant at home. Indeed, families report using Video Mediated Communication (VMC) services like Skype and FaceTime to help their children develop and maintain relationships with parents who are separated from them by work (Yarosh & Abowd, 2011), divorce (Yarosh, Chieh, & Abowd, 2009), immigration (Madianou & Miller, 2012), or military deployment (Yeary, Zoll, & Reschke, 2012), as well as with remote grandparents (Ames, Go, Kaye, & Spasojevic, 2010).

The accessibility of VMC technologies may be especially important for children under 7 years of age, as they tend to have difficulty using audio-only media like telephones to communicate (Ballagas, Kaye, Ames, Go, & Raffle, 2009). Furthermore, because the use of audio-only telephones requires verbal and cognitive skills that they have not yet acquired, infants and toddlers under 2 years of age are especially unlikely to be able to use such media effectively. VMC may offer an alternative, but little is known about its use by or effects on children.
Preliminary studies have shown that toddlers remain content for longer when they have access to a parent via VMC than when they are completely alone (Tarasuik, Galligan, & Kaufman, 2011) or when they have access to a parent via audio-only telephone (Tarasuik, Galligan, & Kaufman, 2013). However, there is also evidence that two-dimensional media can be difficult to process for toddlers (for review see Barr, 2013), which may suggest that VMC technologies would make little difference to such young children. Even as the prevalence of smartphone ownership among families with young children has grown in the past 5 years – 75% of families with children between 0 and 8 years of age own smartphones or some other mobile touchscreen device (Rideout, 2013) – and has made video chatting more accessible, there is still very little data on the frequency of VMC usage among these young children.

While VMC, as a visual medium, has the potential to offer a more effective communication alternative for toddlers, parents may face a dilemma when deciding whether to use this technology to support their children’s long-distance relationships. The American Academy of Pediatrics (AAP) recommends that children under 2 years of age avoid screen media exposure entirely, due to the absence of evidence supporting its benefits and the potential for negative effects (American Academy of Pediatrics, 2011). Public recommendations like these may lead parents of infants and toddlers to avoid using VMC, simply because it is a type of media exposure.

Given the AAP’s recommendation on the one hand, and the potential benefits of VMC on the other, are families actually using this highly accessible technology with their toddlers at home? While nationally representative surveys report that those infants and toddlers who are exposed to screens spend an average of at least 1.5 hours per day with them (Rideout, 2011),

---

1 As stated in the review, several techniques have been demonstrated to mitigate this deficit, including repeated viewings of the video demonstration and also the reduction of the demonstrated task’s difficulty (Barr & Hayne, 1999). Generally, this difficulty decreases gradually by age and task difficulty until around 3 years.
none of these have yet collected data on VMC usage among such young children. This survey study aims to establish a preliminary understanding of the frequency of VMC usage among families with children under two years of age.

**Method**

An online survey was distributed to families in the DC metro area who had at least one child between the ages of 6 and 24 months. These families were recruited at community events, through parenting listservs, flyers, and through word of mouth. The 38 item survey contained 16 questions regarding the family’s income and parental education, race and ethnicity, the age of person completing the survey, the child’s number of siblings, and other questions about the child’s development; and 13 questions on the family’s and child’s general media usage (questions modeled on Rideout, 2013). An additional 9 items regarding the child’s use of VMC were asked only to those who had ever used this specific technology. All questions about the child’s media usage were asked regarding the day prior to taking the survey in order to minimize memory biases, and participants were instructed to complete the survey with their 6- to 24-month-old child in mind.

Responses from 183 eligible families were collected, and participants were predominantly middle to upper-class (median household income between $100,000 and $150,000 per year), well-educated (79% master’s degree or higher), Caucasian (93% White) mothers (3% fathers) with a mean age of 34 years ($SD = 4.6$). Children from each month between 6 and 24 months were represented in the sample ($M = 13$ months, $SD = 5.5$), 51% of them were female, and 33% had an older sibling.
Results

General Media Exposure

Among these families, 100% owned a mobile phone of some kind and 97% owned a smartphone with both a touchscreen and internet access. Only one family in the sample did not own any devices that could potentially connect to the internet, including a smartphone, tablet, or a laptop or desktop computer. Because free VMC software like Skype and FaceTime is available for all smartphones and internet-enabled tablets or computers, all families but one in this sample could potentially access VMC technologies.

The media exposure data for this sample are reported in Table 1. Statistics from previous nationally-representative media usage surveys for children 2 years and under are presented for comparison purposes. While both this survey and previous ones asked about the child’s media use on the day prior to taking the survey, comparisons across surveys should be made with caution: the current sample was not nationally representative, and important methodological differences exist between this study and those. However, the percentage of children who had ever watched television was similar in our sample to national estimates. The estimated time spent viewing TV and DVDs daily was much lower for our sample; however, the children in the current sample are much more likely to have been exposed to computers (18% in this sample versus 10% nationally) and mobile devices (56% in this sample versus 38% nationally) (Rideout, 2013). Additionally, 29% of parents in this sample reported that their toddler had watched online videos, TV shows, or movies on a mobile device; nationally, 47% of children 0-8 years have watched videos on a mobile device, and 38% have watched TV or movies on one (Rideout, 2013). Not surprisingly, given the high-income status of the present sample, smartphone ownership in this group (95%) was higher than the 2013 national average (63%) of families with
children 0-8 years (Rideout, 2013). However, neither statistic (mobile phone ownership nor video/TV/movie-watching on mobile devices) was broken down by specific ages in previous surveys, so the comparison is not exact.

Table 1
Media use for children up to 24 months, over time.

<table>
<thead>
<tr>
<th>Surveys</th>
<th>2005*</th>
<th>2011**</th>
<th>2013***</th>
<th>This Sample (2014)****</th>
</tr>
</thead>
<tbody>
<tr>
<td>Have ever watched TV (%)</td>
<td>79%</td>
<td>75%</td>
<td>66%</td>
<td>63%</td>
</tr>
<tr>
<td>Have ever watched DVDs/videos (%)</td>
<td>65%</td>
<td>62%</td>
<td>46%</td>
<td>32%</td>
</tr>
<tr>
<td>Have ever used a mobile device (%)</td>
<td>--</td>
<td>10%</td>
<td>38%</td>
<td>56%</td>
</tr>
<tr>
<td>Have ever used a computer (%)</td>
<td>--</td>
<td>4%</td>
<td>10%</td>
<td>18%</td>
</tr>
</tbody>
</table>

Average time spent in a typical day, in hours and minutes (hh:mm)

| Watching TV per day (among all) | 00:34 | 00:42 | 00:44 | 00:16 |
| Watching DVD/videos (among all) | 00:13 | 00:19 | 00:11 | 00:03 |
| Watching TV (among those who watched) | 01:02 | 01:30 | --    | 00:47 |
| Watching DVDs/videos (among those who watched) | ++    | 01:16 | --    | 00:35 |

* Rideout & Hamel, 2006: 6-23 months
** Rideout, 2011: 6-23 months
*** Rideout, 2013: 0-23 months
**** This Sample: 6-24 months
++ Sample size too small for reliable results
-- Question not asked / Data not reported

The sample was examined for latent variables that might distinguish discrete groups of children defined by their usage of various digital devices. Participants were asked whether the child had ever used a/an: 1) educational game device, 2) handheld video game, 3) MP3 player without a video screen, 4) MP3 player with a video screen, 5) cell phone without a touch screen, 6) cell phone with a touch screen, 7) landline phone, 8) tablet device, 9) computer. Frequencies of affirmative responses for each device are reported in Table 2. Given that no parents indicated their child had used a handheld video game device, this question was dropped from the analysis. A latent class analysis revealed that a two-class solution offered the best combination of fit and parsimony: the absolute fit of the model was good ($G^2 (235) = 83.6, p = 1$), and a series of Lo-Mendell-Rubin likelihood ratio tests indicated that, while a two-class solution was a significantly better fit to the data than a one-class solution ($\chi^2 = 148.1, p < .001$), a three-class solution was
not a better fit than the two-class solution ($\chi^2 = 24.1, p = .306$). Finally, while the AIC and BIC decreased sharply from the one-class to the two-class solution (AIC: 1134.5 to 1003.2 and BIC: 1162.2 to 1057.8, respectively), a comparable decrease was not observed for the three-class solution (AIC = 996.5, BIC = 1080.0).

Table 2

<table>
<thead>
<tr>
<th>Device</th>
<th>Affirmative</th>
<th>Probability of yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell phone with a touch screen</td>
<td>87 (47.5%)</td>
<td>.073</td>
</tr>
<tr>
<td>Tablet device</td>
<td>72 (39.3%)</td>
<td>.077</td>
</tr>
<tr>
<td>Computer</td>
<td>33 (18.0%)</td>
<td>.013</td>
</tr>
<tr>
<td>Landline phone</td>
<td>22 (12.0%)</td>
<td>.246</td>
</tr>
<tr>
<td>Cell phone without a touch screen</td>
<td>16 (8.7%)</td>
<td>.046</td>
</tr>
<tr>
<td>Educational game device</td>
<td>12 (6.6%)</td>
<td>0</td>
</tr>
<tr>
<td>MP3 player without video screen</td>
<td>9 (4.9%)</td>
<td>0</td>
</tr>
<tr>
<td>MP3 player with video screen</td>
<td>5 (2.7%)</td>
<td>0</td>
</tr>
<tr>
<td>Handheld video game</td>
<td>0 (0%)</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 2 reports the parameter estimates for the 2-class solution, including: 1) the proportion of the sample assigned to each class and 2) the probability of an affirmative response among members of the sample assigned to that class. From these results, it appears that there are two distinct latent groups of individuals: those whose children do not use devices (class 1: low usage), and those whose children do (class 2: high usage). The posterior probabilities indicated a high degree of certainty in the assignment to individuals into class 1 (posterior probability = 96.9%) and class 2 (90.3%), and overall model entropy was acceptable (.80). On average, children in the high device usage group were exposed to over twice as many minutes of media - 36.2 minutes (SD = 54.3) - as children in the low device usage group (class 1; $M = 15.1, SD = 31.3$), a statistically significant difference ($t (174) = 3.09, p = .002$).
VMC and Phone Use

According to this survey, VMC is the primary method of remote communication among infants and toddlers. Families in this sample reported using VMC with their infants often. In fact, 85% of survey participants reported having ever used it with their infants, while 60% reported using it at least several times a month and 37% reported using it regularly at least once a week. Thus, those who use VMC with their infants tend to use it quite often. VMC usage also remained equally high across both the younger and older infants (84% of 6- to 16-month-olds, and 88% of 17- to 24-month-olds).

Families that use VMC also appear to prefer this method of remote communication over other methods available to them: Of those who used VMC, only 43% reported that their young child ever used voice-only telephone calls to communicate with others. As might be expected, the extent of this preference differed significantly by the child’s age, such that only 34% of 6- to 16-month-old VMC users had ever used audio-only phone calls, while 63% of 17- to 24-month-olds did ($\chi^2 = 12.64, p = 0.001$).

An analysis was conducted to examine whether membership in the low or high device usage classes (See previous section) affected VMC exposure. While class membership was not related to child gender ($p = 0.306$) or parental education ($p = 0.969$), it was associated with whether children had used VMC ($\chi^2 (1) = 6.62, p = .010$). A smaller proportion of parents who reported low device usage (78.2%) said that their children had used VMC than of parents who reported high device usage (91.7%). It is of note that a high proportion of parents whose children did not otherwise use digital devices reported that their child had used VMC.
Experiences with VMC

Families who reported using VMC with their toddlers were also asked a series of questions regarding the typical participants in the interactions, what methods were most often used to conduct the calls, and their motivations for using VMC.

**VMC usage patterns.** VMC was used far more often to communicate with individuals living 30 or more miles away (91%) than with those nearby (9%), ($\chi^2 = 104.06, p < 0.001$). In approximately three out of four cases (76%), mothers were reported as being physically present most often with the child during these interactions. Children interacted typically with two individuals on the screen at a time (55%, the mode), usually with grandparents (85%).

Nearly all participants reported using either FaceTime (48%) or Skype (41%) most often to conduct the video calls, and the most popular devices to use were mobile phones (39%), tablets (32%), and laptops (26%). Twenty-six percent of the families who used VMC reported that they had ever used an open video connection – when the video link is left open for an extended period with no particular pressure to talk or remain in view on camera.

**Open-ended questions.** Several topics in the survey were assessed through open-ended questions. First, families’ motivations for using VMC were evaluated with a question asking them to list their top two reasons for using VMC with their child. All 155 individuals who reported using VMC with their child provided at least one reason, and 123 provided a second, for a total of 278 reasons. Four hundred twenty-one unique words were used in the responses, but some words and phrases were used more often than others. To visualize the data, a word cloud was created and is shown in Figure 1. The words parents used most often to describe their motivations were consistent with the survey data generally, and 92% of them fell into at least one of three categories. The first category encompassed the people with whom these families used
VMC, and these were mentioned in 88% of the reasons (266 instances, or 13% of all words used). For example, words for generic family members (e.g. “family” and “relatives”) were used 128 times; variations on the word “grandparent” were used 93 times; words for parents came next (e.g. mom, mother, dad, father, parent) with 32 instances; and “friend” or “friends” was used 13 times. The second category included words that describe types of interaction, and these were mentioned in 68% of reasons (213 instances of interaction words, or 10% of all words used), with visual words leading the group (e.g. “see,” 89 instances) above words like “talk” (45 instances) and “communicate” (42 instances). The third category consisted of adjectives related to the distant location of the screen partner and were used in 44% of the reasons (112 instances, or 5% of all words used), with variations of the phrase “live far away” used 84 times. It is not surprising that these categories would appear, since communicating with, and especially seeing, far away individuals is the express purpose of VMC.

Second, participants who reported that they had ever used an open video connection were asked their reason for using it with their child as an open-ended question, and all but one responded (n = 39). Nearly half of these (46%) reported that they used open video connections because it conveniently allowed one participant (usually the parent) to remain in conversation using VMC while another participant (usually the child) was able to come and go, as in the case

![Figure 1. Most frequently used words to describe family motivations for using VMC](image-url)
of playing off-screen. For example, one parent wrote, “[My] child will say hi for a moment, then grandparents watch him run around and/or play. [My] child will check back in occasionally…[with] no pressure for him to sit on my lap to interact constantly.” Similarly, parents also reported that open video connections were simply convenient or less stressful (23%). For example, one mother wrote, “It’s just very laid back…. Having less focused conversations is less stressful, and I am willing to do it more often.” Some (13%) also reported that open video connections helped them feel that the remote relative was “in the same room,” that the interaction was “more real,” or that it made it feel “as if they were with us.” These comments related to the enjoyment families received from maintaining routines and replicating more casual family interactions; for instance, one parent wrote, “[We can] chat and socialize with family as if they were with us on a weekend morning.” Another parent wrote, “It feels more like a ‘real’ interaction. Grandparents can see him running around and playing in real time. He can come over and say hi if he wants, just as if they were in the room with him.” One parent also reported using an open video connection to allow relatives to be present for important childhood events: “[We use it to] allow grandparents to experience parts of [our child’s] life (i.e. Christmas morning).” This VMC method appears to have resonated positively with the parents who use it, and their reports suggest that it may have potential for increasing the social presence of relatives and the naturalness of interactions.

Third, participants had the opportunity to provide additional comments about their child’s media use at the end of the survey. One hundred two individuals provided additional information in response to this open-ended question. Consistent with the latent class analysis, many of them (38%) spontaneously reported that they placed constraints on media usage with their infants and toddlers; an additional 14% reported that the child was very rarely exposed to
media, but didn’t specify a particular restriction. A typical example of a media restriction comment is, “We limit her video exposure as she's under age 2.” These comments suggest that many parents in the sample were probably aware of the AAP’s recommendation that media exposure be avoided for children under the age of two. Such comments about parental media restrictions were rarely (33%) written without caveats, however. Thirty-three percent of them conveyed media restrictions while also describing the child’s incidental or unintentional background media exposure. For example, typical comments included, “We do not let [our child] watch TV but do sometimes have on a sporting event in the background like golf.” Thirty-six percent of these comments conveyed media restrictions with an intentional exception made for VMC usage, in agreement with the class results reported in section 3.3.2. These included, “We don't let our daughter use media except for face timing [sic] with relatives,” and, “We try to be very thoughtful about [our child’s] media intake…Facetime calls with grandparents don't count.” These comments indicate that even those who do restrict their infants’ and toddlers’ media usage may see VMC as a worthwhile exception to their otherwise strict adherence to the AAP’s recommendation. Similarly, over half of the comments clarifying the child’s minimal media exposure also included exceptions for VMC, like, “[My child] hasn't used any media yet, just Facetime [sic].” It appears, then, that at least some parents are introducing VMC to their infants earlier or more often than other types of media. It is important to note that these comments were made spontaneously in response to an open-ended question about the child’s general media use, however. To explore this potentially common parental attitude further, targeted questions about this topic should be asked in future studies.
Discussion

This survey demonstrates that the use of VMC among infants and toddlers under two is remarkably common, even among those whose parents know and otherwise follow the AAP’s media usage guidelines.

It also provides a novel insight into both early childhood media exposure and the modern family media environment. It became clear from this survey that VMC is being used primarily for communication with far away screen partners. This suggests that it is not being used to replace face-to-face interactions with local friends and family; instead, it is being used to allow the child to interact with remote relatives with whom they may not have frequent in-person contact. Strong family ties are important for healthy child development (Shonkoff & Phillips, 2000), and this technology may offer a needed resource for preverbal children to maintain those relationships with their relatives. Recent research has suggested that the visual availability of an absent parent on screen can provide comfort to young children (Tarasuk, Galligan, & Kaufman, 2011) and reduce a child’s anxiety about parental separation (Yeary, Zoll, & Reschke, 2012), and it’s possible these results would extend to other relatives as well. Thus, these findings have important implications for those who are crafting policies for media use with young children and for understanding changing patterns of communication among families that are geographically separated.

Media policy makers should also pay careful attention to the plethora of new and mobile devices that are widely available across socio-economic groups (Rideout, 2013). In this sample specifically, parents reported that young children spent less time watching videos and DVDs than reported in previous studies, but given their very young age, a surprising number reported exposure to videos via mobile devices. This may suggest that TV- and video-watching is now
occurring more often on mobile devices and less often on traditional TV screens. This is consistent with the downward trend across studies in recent years in the number of young children who have ever watched traditional DVDs or videos (see Table 1). Also, in the present sample patterns of phone usage, but not VMC usage, differed by age. While VMC usage remained high across the entire sample, among VMC users the percentage of those who had also used telephones to communicate with others increased for children over 17 months. This is not surprising, given that children typically experience a rapid increase in vocabulary between 1.5 and 2 years of age (Fenson et al., 1994). While non-visual communication technologies remain a problematic mode of communication through 7 years of age (Ballagas et al., 2009), this early communicative milestone may make it easier, or at least more likely, for these toddlers to utilize audio-only telephones.

Furthermore, parents who placed restrictions on, or whose child otherwise had minimal exposure to, other forms of media exempted VMC from their limited screen use. The AAP should keep this parental perspective regarding VMC in mind when advising families about infant media exposure, particularly given that the recommendation is based primarily on research conducted on television usage. While this survey did not assess the average length of families’ VMC calls, our own observational research (McClure, Chentsova-Dutton, Holochwost, Parrott, & Barr, under review) indicates that families with children under two spend an average of 20 minutes in each VMC call. This is a significant addition to existing estimates of childhood media exposure times, particularly for those (nearly 40% of this sample) who use VMC at least once per week. Future nationally representative surveys of childhood media use should include questions about VMC usage to more accurately estimate the time per day children spend with screens.
There are some limitations to these findings. This survey was not intended to be nationally representative, so one should use caution when generalizing the results. The survey does provide a fairly representative look at the media usage patterns among highly-resourced families (both high education and high income). There is, however, reason to believe that VMC is being used nationally too. VMC usage was well spread across income groups in the sample: For all income levels, at least 50% used video chat at least several times a month. Furthermore, 75% of families nationally with children between zero and eight years own smartphones or some other mobile touchscreen device (Rideout, 2013). These devices have made video chatting easily accessible to families, so determining its usage among young children is an important endeavor. This survey is a first foray into estimating video chat usage among infants and toddlers, and the results should encourage the creators of nationally representative media usage surveys to include questions about VMC in future studies.

**Conclusion**

VMC surfaced in this study as a substantial and previously unreported source of media exposure for children under 2 years of age. There are three key findings from this timely survey. First, VMC is widely used by children under 2 years and is categorized differently by parents from other types of screen media. Second, parents commonly use VMC with their infants and toddlers to maintain connections with geographically remote family members. Third, the media landscape is changing rapidly and so traditional forms of media viewing and communication need to be revisited by both researchers and policy makers.

These findings make it clear that more research is required on VMC usage among children in this age range. As a preverbal population, the visual characteristics of VMC make it an appealing resource for remote communication; however, these children are also in a very
sensitive period of development, during which media exposure should be approached with caution. Future research should include not only precise measures of the quantity of infant exposure to this technology, but also the quality of that exposure, as determined through naturalistic observations of families using video chat at home. Controlled laboratory experiments are also called for, to test the degree to which infants are socially and emotionally engaged during video communication, as well as the developmental effects of extended exposure. Parents use VMC to familiarize their infants with remote relatives, but additional research regarding how very young children perceive these interactions is needed.
Chapter III: Bridging the Distance: How babies and remote relatives connect using video chat at home

Portions of this chapter are excerpted from a book chapter submitted as:

Introduction

Many young children today are geographically separated from their immediate or extended families, and remote communication is an important resource for them (Chapter I, this dissertation). Telephones can be difficult to use for children under 7 years of age, however, and video chat has been suggested as a viable alternative (Ballagas, Kaya, Ames, Go, & Raffle, 2009). While families with children under two are now using video chat to help their babies and toddlers keep in touch with remote relatives (Chapter I, this dissertation), there is no prior research studying natural video chat use by babies under two in their own homes. It is therefore unclear whether babies are capable of using these technologies successfully to interact with others.

In particular, it has not been demonstrated whether babies in this age range can and do pay attention to extended video mediated interactions. Babies are still developing the skills associated with social attention, and the challenges presented by video chat may make these skills more difficult to access. This chapter is broken down by the major developing skillsets that are likely to be challenged by video chat interactions, followed by the barriers to using those skills successfully in the video chat context.

Typical Development of Associated Skills

Children under two years of age are still in an early period of attention development. According to Colombo (2001), the (non-volitional) attainment of attention, or simple alertness, is
possible by full-term, even within the womb, although newborns spend little time in this state. Spatial orienting, the orienting of attention to specific objects in the environment, also begins very early in the newborn’s life and begins developing rapidly at 3 months. Attention to object features, like color and texture, conceived as a single form, is established by about 4-5 months of age, and by at least 6 months babies’ attention is drawn more to novel objects than to ones they’ve seen before. Finally, the apparently voluntary maintenance of attention begins to appear between 4 and 6 months of age. All of these nascent attention systems continue to develop across the first two years of life. In particular, the very challenging task of maintaining attention continues to develop throughout childhood (Betts, Mckay, Maruff, & Anderson, 2006). Of course, there are many strategies that adults can and do employ to help attain and maintain babies’ and toddlers’ burgeoning attention – like touching the child, using highly contingent interaction styles, overtly directing their attention using gaze and pointing (i.e. joint visual attention), and engaging the child in stimulating play activities (i.e. joint pretend play).

**Joint visual attention.** *Joint visual attention* (JVA), or “following the direction of attention of another person to the object of their attention” (Butterworth, 2004, p. 213), may be difficult for children to accomplish successfully in the context of video chat. JVA, the most overt way of attaining babies’ visual attention, has been said to form the foundation of shared social moments (Butterworth, 2004), so compromises in JVA have the potential to disrupt the interaction in critical ways for both the baby and the remote relative.

JVA is still in the process of formation for children under 2 years of age. Before 9 months, adults tend to adjust their own gaze to the attention of the infant, which is called “supported” JVA (Bakeman and Adamson, 1984). Starting around 9 months, the infant begins to more frequently initiate the exchange of attention, to reference objects, and to follow the adult’s
gaze, which is called “coordinated” JVA (Bakeman and Adamson, 1984). At 12 months, babies still fail to search for objects behind them when an adult’s visual attention is directed there (Butterworth & Jarrett, 1991) – a behavior that only emerges around the age of 18 months (Butterworth & Cochran, 1980). Furthermore, it is also not until infants are 18 months old that they begin searching the area beyond their peripheral vision when their attention is directed there using JVA.

**Joint pretend play.** Babies under two years of age are only beginning to develop their pretend play skills. These skills increase rapidly between 15 and 18 months of age and continues to emerge throughout the second year of life (Piaget, 1962; Lillard, Pinkham, & Smith, 2010), the upper end of the age range considered here.

**Barriers During Video Chat**

The typical strategies adults use to attain and maintain attention may prove challenging for remote relatives in the context of video chat. For example, physical contact, an important resource for maintaining attention in infants (Stack & Muir, 1990), is non-existent between the remote relative and the baby. Social contingency, the “sequentially dependent, close temporal relationship” between an adult’s response to an infant’s behavior (Dunham & Dunham, 1995, p. 160), and reciprocity, the sensitivity and appropriateness of their responses (Dunham & Dunham, 1995), can be disrupted due to audio-visual delays associated with bandwidth limitations or other technological problems. Given the importance of these behaviors for the maintenance of infant attention (Hains & Muir, 1996) and for forming social preferences (Bigelow, 1998), such disruptions are important to consider. Furthermore, the aforementioned JVA and JPP skills may be difficult for babies and their caregivers to transfer successfully into the video chat context.
Joint visual attention. The limitations of babies’ JVA skills at these ages may make it difficult for them to successfully participate in JVA during video chat interactions. First, consider that pre-installed or manually attached cameras used for video chat tend to be located above the screen, on which the remote individual and their environment are displayed. When a user attempts to make eye contact with their social partner or look at something in their environment, then, the user is directing his or her gaze several inches below the camera’s aperture. The image that is transmitted to the user’s partner is that of the user gazing several inches below the receiver’s eye level or below the item referenced in their environment (Grayson & Monk, 2003). This misalignment of gaze could make both supported and coordinated JVA between infants and their remote relatives more difficult.

Furthermore, the shared visual space between infants and their video mediated social partners is limited to what can be displayed through their monitors. Given the position of a webcam relative to an infant participating in a video call, if mediated social partners reference something in the infant’s environment, the object is likely to be behind the infant, beyond the baby’s peripheral vision. For this reason, video mediated JVA may be limited for children under 18 months without the assistance of a physically present adult. Moreover, given the constricted and disconnected “shared” visual environment in VMC, it is conceivable that video mediated JVA toward objects behind the toddler will be delayed relative to the face-to-face emergence of this ability. It is possible, then, that many children could continue to struggle with adult-initiated, video mediated JVA through their second birthdays.

Babies and toddlers may still be able to participate in successful video mediated JVA, however, if the adults involved adapt to their needs and to the limitations of video chat. A baby may be able to follow the gaze of a remote relative whose attention is directed toward an object
or event within the relative’s own environment (Hood et al., 1998) – that is, so long as the object or event is within the view of the relative’s camera. For example, if the adult partner brings a tea cup up to the camera to show the child, the adult has bypassed the limitations of the camera’s location, allowing the child to more easily follow the adult’s attention. Babies may also be able to initiate this “within-screen” JVA, directing their remote relative’s attention to items nearby in their own (i.e. the babies’) environment. This is an encouraging possibility, and one that may provide the opportunity for both generally successful social interaction and attempts at triadic (i.e. baby – relative – object) play across the medium.

Joint pretend play. It may be difficult for toddlers to transfer their developing pretend play skills across a screen interface. Pierroutsakos and Troseth (2003) demonstrated that babies’ understanding of the representational nature of screens also develops gradually across the first 2 years. At 9 months, babies still grab at objects on the screen and try to manipulate them, behaviors that decrease by 15 and 19 months. It’s possible, then, that only babies nearing the end of their second year will participate successfully in cross-screen pretend activities. Parents regularly scaffold and engage in pretend play with their babies (Lillard et al., 2010), so their participation may play a special role in this context.

The current study

How are families responding to these barriers to attaining and maintaining babies’ attention when using video chat? The current study utilized an observational design to investigate families’ strategies for overcoming such barriers when using video chat naturally at home. The results are broken into five sections. First, the usage of screen-mediated JVA is examined, with specific hypotheses in mind. All babies were expected to initiate the more accessible “within-screen” JVA more often than “across-screen” JVA, with older babies being
more likely to begin initiating “across-screen” JVA as well. It was expected that present adults would assist in JVA episodes, as interpreters of misaligned gaze and point cues, and that this assistance would improve the success rate of JVA instances involving the baby.

Second, we describe the occurrence of cross-screen JPP. Prior literature does not allow us to predict how often JPP would be attempted, nor whether these attempts would be successful, particularly among the younger children. This study was meant to address this gap. Third, the use of physical contact by the present family members is briefly explored, with a descriptive analysis of touch and affection in the physical absence of the remote relative. Fourth, the frequency and duration of technical problems that compromised the contingency of the interaction are documented.

Fifth, behaviors are tested for their predictive power of the child’s overall attention to the video call. Given the early stage of these babies’ attentional development, it was expected that external scaffolds would play a role in supporting the child’s attention during the call period. For this reason, it was anticipated that the use of across-screen JVA would be a significant predictor of the child’s attention to the screen. Furthermore, given the well-studied relationship between a child’s attention and the adult’s contingency and reciprocity (e.g. Hains & Muir, 1996), the objectively-rated sensitivity² of the remote relative was also expected to play a role in predicting the child’s attention. Exploratory analyses of other relationships with attention were conducted without specific hypotheses for this context.

---
² Assessed using a scale that incorporates the adult’s contingent and appropriate responsiveness to the child (see section Behavioral Coding, below).
Method

Participants

Families with a child between the ages of 6 and 24 months were recruited from the D.C.-metro area via flyers, listservs, community events, and word of mouth. All potential families were sent an electronic media usage survey, and only families who reported using video chat with their babies at least once a month were selected for the study.

Of the 28 observed video calls, 25 were made to remote grandparents. The remaining three were either made to the child’s father ($n = 2$), who was at work, or to the child’s aunt ($n = 1$). To maintain sample homogeneity\(^3\), these three calls were dropped from the remaining analyses, with a final sample of 25 video calls, all of which took place with remote grandparents (56% both grandparents, 36% grandmothers alone, 8% grandfathers alone).

Babies had a mean age of 16 months ($SD = 5.27$), and just under half were female (48%). Mothers were always physically present during the calls, and fathers joined them in half of the sessions (52%).

Procedure

A semi-naturalistic observational method was employed for this study. Two researchers went to each family’s home when the family was already planning to use video chat with their child. Families were instructed to interact as they typically would during a video chat session with their remote relative. The researchers recorded the interaction using small, hand-held video cameras while standing unobtrusively outside the interaction space. One researcher recorded the screen, while the other researcher recorded the physically present family members. Afterwards,

\(^3\) These calls were statistical outliers on one or more of the following characteristics: call length, time since seeing the relative in-person, and adult-centered interaction style.
the two video recordings of the calls were synced side-by-side and coded for behaviors of interest.

**Behavioral Coding**

Two coders were trained to identify behaviors of interest using pilot videos for the study. Both then coded 21% of the recordings for reliability purposes, and inter-rater reliability ranged between ICC = 0.70 and 1.00. Any behavior that did not appear in at least six calls was dropped from all further analyses. After reliability was reached, each coder was responsible for coding 50% of the recorded video calls. Calls that were longer than the average call length were coded for approximately 20 minutes of their duration.

The total frequency of each behavior per call was calculated. Then variables were created to demonstrate the occurrence of specific subcategories within each code, as a ratio between the total number of subcategory instances to the total number of instances of the general code per call. For example, the construct for “JVA success rate” is the ratio of the number of successful JVA instances to the total number of JVA instances that occurred in the call; and the construct for “parental involvement in device touches” is the ratio of the number of times a parent attempted to touch the device to the total number of device touch attempts during the call. These ratios were converted to percentages for ease of interpretation.

The code definitions and their corresponding reliability coefficients are presented in Appendix A. An abbreviated set of definitions is provided here as well.

**Joint Visual Attention (JVA).** JVA was defined as when one individual directed another individual’s attention to a third object, person, or event. Success was defined as when the person whose attention is being directed directs their attention to the object or person intended by the initiator. There were four types of JVA coded (See Figure 2):
• **Within-screen JVA.** The JVA initiator directs a partner’s attention to an object or person on their own side of the screen (e.g. a baby brings a toy to the screen to show the grandparent).

• **Across-screen JVA.** The JVA initiator directs a partner’s attention to an object or person on the other side of the screen (e.g. the baby points to her grandmother’s dog).

• **Within-beyond screen JVA.** The JVA initiator directs a partner’s attention to an object or person on their own side of the screen that is beyond the view of their partner (e.g. the baby points out the window to a passing truck, with the grandparent cannot see).

• **Across-beyond screen JVA.** The JVA initiator directs a partner’s attention to an object or person on the other side of the screen that is beyond their own view (e.g. the baby asks to see the grandparent’s dog, who is currently off-screen).

![Figure 2](image1.png)  
*Figure 2.* Three types of video mediated JVA. Panel (a): a grandmother shows the baby an apple, initiating within-screen JVA. Panel (b): a baby points at an object on the laptop screen, initiating across-screen JVA; Panel (c): a baby points to an object beyond the view of the camera, initiating beyond-screen JVA. Within-beyond screen JVA not pictured, due to its verbal nature.

**Joint Pretend Play (JPP).** JPP was defined as when the participants engage in a social activity that involved imaginary or impossible objects, actions, or ideas. This included trying to pass objects through the screen. Success was defined as when the individual invited to play
reciprocated by entering into the shared play activity (e.g. the grandmother accepts the pretend cup of tea from her grandchild and pretends to drink it). There were two types of JPP coded:

- **Within-screen JPP.** Joint Pretend Play between two or more individuals that takes place on only one side of the screen. The Play initiator is on the same side of the screen as the individual invited to play (e.g. the baby pretends to share a cup of tea with her mother).

- **Across-screen JPP.** Joint Pretend Play between two or more individuals that takes place across the screen. The Play initiator is not on the same side of the screen as the individual invited to play (e.g. the baby pretends to share a cup of tea with her remote grandmother).

**Child Attention.** The child’s attention was defined as the amount of time the child spent in screen-directed behaviors (e.g. gazing, gesticulating, vocalizing, showing things, or participating in activities facing toward the device) in each call. The number of times the child turned his or her attention away from the call and then returned his or her attention back to it was also noted.

**Displays of physical affection.** Displays of affection were defined as active, non-habitual displays of physical affection between partners (e.g. kisses, hugs). All utilitarian touches were excluded.

**Attempts to touch the device.** Attempts to touch the device were defined as instants in which a device touch (or touch attempt) is made, including when a device changes hands or is adjusted. Success was defined as whether the individual made contact with the device. Verbal encouragements of the attempt were defined as when an individual other than the one making the attempt verbally encouraged the attempt (e.g. “Give Grandma a hug!” and “Can you press the ‘end call’ button?”). Verbal discouragement of the attempt was defined as when an individual other than the one making the attempt verbally discouraged the attempt (e.g. “No, don’t touch!"
You’ll hang up on her!”). Finally, a neutral attitude toward the attempt was defined as when neither a verbal encouragement nor discouragement of the touch attempt was present.

**Technological Problems.** Technological problems were defined as problems with the technology that involved significant visual, audio, or visual-audio delay; this included problems leading to call termination. Durations ended when the problem was completely resolved, either within the same call period or at the restart of a dropped call. The number and source of the problems (e.g. the technology itself, the baby, the mother) were also coded.

**Technological Talk.** Technological talk was defined as durations when participants talked about the technology they were using for the video call, including talking about any technical difficulties (e.g. “This connection is terrible,” and “How do I get my video to work?”). One sub-category was also coded:

- **Technological talk about seeing.** Technological talk regarding any individual’s ability or inability to see a partner or partners through the camera’s view (e.g. “Can he see me?” and “We only see part of your face!”). This includes requests to adjust the camera’s position or angle.

**Center of Interaction.** The center of the interaction was coded using a scale of 1 to 3, rating whether the call was: (1) mostly child-centered; (2) equally child-centered and adult-centered; or (3) mostly adult-centered.

**Interaction- or Observation-Based Call.** A scale of 1 to 3 was used, rating whether the call was based: (1) mostly in interaction with the child; (2) equally in interaction with and observation of the child; or (3) mostly in observation of the child.

**Baby Activity Level.** The activity level of the child was defined as how often the child moved from place to place during the call. A scale of 1 to 4 was used to rate the child’s activity
level as: (1) Stationary (i.e. the child does not change locations at all); (2) Low Activity (i.e. the child changes locations very rarely); (3) Moderate Activity (i.e. the child changes locations regularly); or (4) High Activity (i.e. the child changes locations nearly constantly).

**Device Movement.** The frequency with which the device was moved or transferred throughout the interaction was rated on a scale of 1 to 4 as: (1) Stationary (i.e. the device does not move at all); (2) Low movement (i.e. the device moves very rarely); (3) Moderate movement (i.e. the device moves regularly); (4) High (i.e. the device moves nearly constantly).

**Adult Sensitivity.** The sensitivity of the adult participants in the call was assessed using the *Parental Sensitivity Scale* from *The Emotional Availability Scales, abridged Infancy/Early Childhood Version* (Biringen, Robinson, & Emde, 2000). The scale incorporates a variety of indicators including the adult’s accuracy in reading infant signals, appropriate responsiveness to such signals, adult affect, awareness of timing, and flexibility. Ratings on this 1-9 scale were recorded continuously during the video recordings using a joystick, and the average value for each adult in each call was used for analysis. Coders selected a single primary present adult and a single primary remote adult to score on this measure.

**Results**

**General Call Characteristics**

The average call length was 20 minutes 18 seconds (*SD* = 9 minutes 28 seconds). Babies paid attention for 41% of the call length on average. There was no relationship between the babies’ ages and the call lengths (*r* = 0.25, *p* = 0.23) or the percent of the call during which they paid attention (*r* = - 0.11, *p* = 0.62).

---

4 This does not include time spent restarting dropped calls. Removing two outliers in call length (41 minutes 5 seconds, and 46 minutes 11 seconds), the average call length was 18 minutes 16 seconds.
Laptops were used most often to conduct the video calls (44%; 2 families projected it onto a large television screen), followed by touchscreen tablets (32%) and smartphones (24%). There was no difference in the average age of children using mobile versus non-mobile devices, $t(23) = 0.25, p = 0.81$. Age also played no role in the activity level of the babies and toddlers ($t(7.87) = -1.11, p = 0.30$), with the majority (72%) of children rated as having a moderate to high activity level. The devices were in motion as well (48% low movement), with mobile devices moving from place to place significantly more often (64% moderate to high movement) than laptops (0% moderate to high movement), $\chi^2(1) = 11.05, p = 0.001$. There was no significant difference in infant activity level across mobile and non-mobile devices ($\chi^2(1) = 2.97, p = 0.17$).

Nearly all (88%) of the calls were rated as being mostly child-centered, while 0 were mostly adult-centered, and only 12% were rated as equally child- and adult-centered. There was no difference in the average ages of babies involved in child-centered or equally child- and adult-centered interaction, $t(23) = 0.50, p = 0.62$. Similarly, the majority of calls (72%) were rated as being focused mostly on interaction with the baby (36%) or equally interaction- and observation-based (36%), with the minority (28%) rated as being centered mostly on observation of the child. There was a significant difference in the average age of the babies involved in each kind of call, $F(2, 25) = 4.94, p = 0.02$, such that babies in observation-based calls ($M = 11.60, SD = 1.73$) were significantly younger than babies in interaction-based calls ($M = 18.44, SD = 1.52$), $p = 0.02$; and were marginally younger than those in calls rated as equally interaction- and observation-based ($M = 17.33, SD = 1.52$), $p = 0.06$.

---

5 Because there were zero laptops that moved at a moderate to high level, this result should be read with caution.

6 Given the small number of calls rated as equally child- and adult-centered ($n = 3$), this result should be read cautiously.
Joint Visual Attention

Before hypothesis testing was conducted, descriptive statistics were performed to describe the usage of JVA among the sample. Then inferential statistics were used to test specific hypotheses. Finally, exploratory tests were conducted to examine research questions for which there were no specific hypotheses.

JVA characteristics in video chat. First, the prevalence of JVA usage and its overall success rate across the sample were assessed. All 25 calls included instances of joint visual attention, and the mean number of JVA attempts per minute was 2.4 (SD = 0.97; Min = 1, Max = 4). On average, JVA attempts were fairly successful (84%; SD = 11), and the minimum success rate in any call was 63%. Surprisingly, age was related neither to the amount of total JVA used \(^7\) (r = - 0.13, p = 0.53) nor its success rate (r = 0.20, p = 0.34).

Second, the frequency with which the types of JVA were used was examined. Within-screen JVA, theoretically the most accessible type of video mediated JVA, was the most common type witnessed in the calls, taking up an average of 64% of total JVA instances per call (SD = 14). Furthermore, it was the most-often used type of JVA in 88% of the calls. Across-screen JVA, which is theoretically more challenging, occurred in 33% of total JVA instances per call, on average (SD = 15), and was the most often used type of JVA in only 12% of the calls. Within-beyond screen JVA was so rarely used in the calls (<1%) that coding reliability could not be established. Finally, across-beyond screen JVA occurred in only 3% of JVA instances per call on average (SD = 3).

Third, the degree to which the family members were involved in the JVA instances was surveyed. For the purposes of this chapter, individuals will be called JVA initiators when they were responsible for initiating an instance of JVA; they will be called JVA directees when they

---

\(^7\) This was calculated per minute, to account for differences in call lengths
had their attention directed to a third object by an initiator; and they will simply be called

*participants* when their role is not being specified (e.g. their overall participation, regardless of
their role in it). The average level of participation of each individual in each role is presented in

Table 3.

**Table 3**

*Average participation and role type in video mediated JVA*

<table>
<thead>
<tr>
<th>Role</th>
<th>Total Participation(^8)</th>
<th>Initiations</th>
<th>Being Directed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Babies</td>
<td>87% (SD = 17)</td>
<td>17% (SD = 14)</td>
<td>69% (SD = 18)</td>
</tr>
<tr>
<td>Mothers</td>
<td>55% (SD = 19)</td>
<td>35% (SD = 15)</td>
<td>11% (SD = 9)</td>
</tr>
<tr>
<td>Fathers</td>
<td>11% (SD = 17)</td>
<td>7% (SD = 11)</td>
<td>2% (SD = 4)</td>
</tr>
<tr>
<td>Remote Grandmothers</td>
<td>43% (SD = 20)</td>
<td>30% (SD = 17)</td>
<td>9% (SD = 7)</td>
</tr>
<tr>
<td>Remote Grandfathers</td>
<td>17% (SD = 18)</td>
<td>11% (SD = 12)</td>
<td>4% (SD = 5)</td>
</tr>
</tbody>
</table>

Fourth, the degree to which adults provided assistance in JVA was examined. For example, when a baby initiates within-screen JVA a present adult might help the baby place the object of attention so that it can be seen through the camera. In across-screen JVA initiated by a remote grandparent, an adult might help the baby understand what the grandparent is referencing. Such assistance was provided in an average of 10% of JVA instances across calls (*SD = 8*) and was mostly provided by a present parent (84%, mostly moms), but occasionally by remote grandparents (16%). JVA assistance in each call was unrelated to the baby’s age (*r* = 0.04, *p = 0.85), to the frequency of JVA initiated by the baby (*r* = 0.20, *p = 0.34), or to the frequency of JVA directed toward the baby (*r* = 0.24, *p = 0.25). Beyond the instances of JVA themselves, the adults also verbally checked in with one another to ensure that they could see and be seen. These verbal check-ins discussing the ability (or inability) of the members to see or

\(^8\) Total Participation does not add up to 100% for each individual for two reasons: 1) It also includes instances in which the individual *assisted* in JVA instances (see below); 2) Not all individuals (especially dads) were involved in all the JVA instances. Furthermore, while babies were involved in nearly all the JVA instances, sometimes JVA was exchanged between the adults only or, occasionally, with the baby’s sibling.
be seen through the camera occurred in 52% of the calls, and in 69% of these cases there was more than one check-in during the same call.

Tests of JVA hypotheses. Because within-beyond screen JVA and across-beyond screen JVA were used very infrequently, they were not assessed in the following inferential tests. First, it was predicted that babies would initiate JVA of any kind more often as their ages increased. This hypothesis was marginally confirmed: there was a marginal positive relationship between baby initiations of JVA and the babies’ ages ($r = 0.37$, $p = 0.07$), such that babies initiated more of the total JVA instances as they got older.

Second, it was predicted that when babies of any age initiated JVA, they would be more likely to initiate within-screen JVA than across-screen JVA. This hypothesis was confirmed: The number of babies that initiated at least one instance of JVA-within (80%) was significantly greater than the number who initiated at least one instance of JVA-across (28%), $p<0.001$ (Fisher’s Exact Test).\(^9\)

Third, it was hypothesized that older babies would be more likely to initiate at least one instance of across-screen JVA than younger babies. Using a median split (15 months), a Chi-Square test confirmed that older babies (16-24 months) were more likely to initiate across-screen JVA (46% of them initiated it at least once) than younger babies (6-15 months; 8% of them initiated it at least once), $\chi^2 (1) = 4.42$, $p = 0.04$, $\phi = 0.42$.\(^{10}\) The number of babies in each age group who initiated each type of JVA is presented in Figure 3.

---

\(^9\) Given the small sample size, the value of the McNemar $\chi^2$ test was not calculated in favor of a Fisher’s exact test, which is more conservative and has no test statistic.

\(^{10}\) Due to the very small cell sizes (only one of the twelve younger babies initiated an instance of across-screen JVA), this result should be interpreted with caution.
Fourth, it was hypothesized that parental assistance in baby-initiated JVA would increase its success rate. Parents assisted in JVA so rarely that this hypothesis could not be tested accurately. There was, however, no statistical relationship between parental assistance of JVA and the overall success rate of JVA ($r = 0.12$, $p = 0.58$).

**Exploratory tests.** The remaining analyses regarding JVA were exploratory in nature, so a limited number of questions were pursued and statistical corrections were used where appropriate.

First, we examined whether mothers used JVA as a tool to redirect the child’s attention to the screen specifically. In other words, did mothers initiate across-screen JVA or within-screen JVA more when babies were the directees? A Chi Square test revealed no difference in the number of mothers who initiated across-screen JVA more (56%) or within-screen JVA more (44%) to direct their babies’ attention, $\chi^2 = 0.36$, $p = 0.55$, $\phi = 0.12$.  

*Figure 3.* Number of babies who initiated at least one instance of within-screen and/or across-screen JVA, by age group. There were no older babies who did not initiate at least once, and no babies in either age group initiated exclusively across-screen.
Second, an examination of the relationship between mother-initiated and baby-initiated JVA types was conducted. In other words, which type of JVA do mothers tend to initiate overall, and how does this relate to her child’s preferred JVA type to initiate? Correlation analyses revealed that mothers tended to initiate across-screen JVA more often than within-screen JVA during the calls ($r = 0.50, p = 0.01$), but that baby initiations of across-screen JVA were positively correlated with maternal initiations of across-screen JVA ($r = 0.50, p = 0.02$). In other words, overall, mothers’ and babies’ tend toward initiating different types of JVA (i.e. mother initiations are positively related to across-screen JVA, baby initiations are positively related to within-screen JVA); however, individual mothers’ and babies’ initiations are also related (i.e. if the mother initiates more instances of across-screen JVA, so does her baby; if the mother initiates more instances of within-screen JVA, so does her baby).

Third, were grandparents who were rated as more sensitive more likely to be participants in JVA? And does their sensitivity have any relationship to the success rate of JVA? Neither relationship was significant: The grandparents’ sensitivity levels were unrelated to both their participation rates in JVA ($r = -0.10, p = 0.63$) and the success rate of JVA ($r = 0.33, p = 0.11$).

**Joint Pretend Play**

Hypotheses about cross-screen pretend play were non-specific, so this section is also largely descriptive and exploratory in nature. It was unclear how often JPP would be attempted, nor whether these attempts would be successful, particularly among the younger children.

---

11 Because two parallel tests were used to address this set of questions, the $p$-value used to determine significance was adjusted by the number of tests employed ($p = 0.05 / 2 = 0.025$).
12 For any specific family, within-screen JVA and across-screen JVA usage (as percentages of total JVA) were near-inverses of one another. For this reason, this correlation analysis and others were only conducted using one of the two JVA types.
13 The $p$-value for significance was again set to $p = 0.025$ for these two parallel analyses.
Sixty percent of the calls included at least one instance of joint pretend play. Among calls where it occurred, it occurred at a rate of about 6 instances in a call of average length (20 minutes). Surprisingly, 89% of JPP attempts per call ($SD = 26$) were successful, and the frequency of JPP usage was unrelated to the child’s age ($r = 0.12$, $p = 0.59$). It was also unrelated to the amount of JVA used ($r = -0.20$, $p = 0.35$); however, the percentage of JVA and JPP that were directed toward the baby were significantly associated ($r = 0.58$, $p = 0.02$).

The majority of JPP ($M = 93\%$, $SD = 26$) was shared across the screen. This, of course, means that grandmothers ($M = 76\%$, $SD = 41$) and grandfathers ($M = 38\%$, $SD = 49$) were involved in the majority of the JPP scenarios. These averages, however, are lowered by the grandparents who did not participate in any JPP scenarios during their calls. In actuality, in calls where grandparents were involved in any JPP at all, they were involved in nearly every instance of JPP: Remote grandmothers participated in at least one JPP scenario in 12 (48\%) of the calls, and in these calls the grandmothers were involved in an average of 96\% of the JPP instances ($SD = 12$); Remote grandfathers participated in at least one JPP scenario in 6 (24\%) of the calls, and in these calls the grandfathers were involved in an average of 95\% of the JPP instances ($SD = 12$). In other words, there are two types of grandparents: those who are involved in none of the JPP, and those who are involved in nearly all of the JPP. These types of grandparents were distributed equally across baby age groups, $\chi^2 (1) = 0.03$, $p = 0.87$.

Without specific hypotheses, relationships were explored between the success rate of JPP and the following variables: the child’s age, whether the scenario included assistance from another adult, the relative’s sensitivity, and how often the screen adult was involved.\(^{14}\) Instances of JPP were very rarely unsuccessful, and this was unrelated to the child’s age ($r = 0.32$, $p =\)

\(^{14}\) Because four parallel tests were conducted, the p-value for significance was adjusted to $p = 0.01$. 

38
0.24), assistance provided by other adults ($r = 0.30, p = 0.28$), the relative’s sensitivity ($r = 0.36$, $p = 0.19$), or how often the screen adult was involved ($r = 0.16, p = 0.57$).

**Physical Contact**

There were no specific hypotheses regarding physical contact among the physically present individuals, besides the expectation that there would be plenty of it. For this reason, the analyses conducted in this section were purely descriptive in nature.

Instances of physical affection among the participants were noted in 80% of the calls, with a maximum of 19 instances in a single call. Ninety-one percent ($SD = 23$) of the instances per call involved the baby, and 7% ($SD = 23$) involved a present sibling. Most instances ($M = 87\%$) involved a physically present parent (Moms: $M = 69\%, SD = 38$; Dads: $M = 12\%, SD = 21$).

Families made frequent physical contact with the video call devices as well. Parents were most likely to touch the device: Among all touch attempts in the sample, 60% were by moms and 14% were by dads. Babies made frequent attempts as well and were responsible for 20% of all touch attempts. Babies made such attempts in 80% of the calls. Babies tried to touch the device more often (in attempts per minute) when the device was a mobile device ($M = 0.70, SD = 0.87$) than when it was a non-mobile device ($M = 0.15, SD = 0.19$), $t (14.63) = 2.30, p = 0.04$, Cohen’s $d = 0.87$.

Babies’ attempts to touch the device are not always welcome, however. In fact, 50% of calls in which the baby made at least one touch attempt also included at least one parental verbal discouragement of the baby’s effort. Overall, 28% of baby attempts to touch the device were verbally discouraged. Babies were successful in 73% of their overall attempts to touch the device; however, when the touch was discouraged, they were only successful 43% of the time.
As would be expected, touch discouragements had a highly negative relationship with successful device touches \((r = -0.87, p < 0.001)\). Sometimes parents encouraged infant contact with the device though, as when suggesting and supporting cross-screen affection like “screen kisses” (see Figure 4) between the baby and grandparent. Encouragements were noted in 28% of calls and, overall, 7% of baby touch attempts were encouraged.

![Figure 4. Parents encourage the child to give his grandmother a screen kiss. Left and right panels are simultaneous views of the physically present family (left) and the remote grandparent on screen (right).](image)

**Technical Problems**

Because there were no specific hypotheses regarding technical problems, the majority of this section will be descriptive in nature.

Fifty-four percent of calls experienced at least one technical problem, defined as any significant audio and/or visual delay or disruption, or an unintended termination of the call.\(^{15}\) Twenty-five percent of calls were interrupted by only one significant problem, and 29% were interrupted by two or more, with a maximum of ten problems in a single call. Five calls (21%) were ended at least once by a technical problem; of these, three were restarted once, one was

\(^{15}\) These only include problems that occurred after the calls began (i.e. either visual, audio, or both were present); thus, problems starting the call for the first time were not included.
restarted twice, and one was restarted four times. For the most part – with one outlier – when a call was dropped, it was simply restarted, with a loss of only a minute or less.

One outlier family experienced a technical problem that took 13 minutes 41 seconds to resolve and was dropped from the remaining analyses of technical problems. Among the calls that were disrupted by technical problems, the average duration of the problems was 43 seconds ($SD = 41$ seconds).

Of the 45 total instances of technical problems in the sample, 67% were caused by the technology itself (e.g. poor connections). Twenty-four percent were caused by the baby, as when babies hang up the phone by touching the screen of a touchscreen device. In 4% of cases, a grandparent was responsible for the problem – for example, when a grandparent was unsure how to start the audio or video input for the call. The remaining 4% of problems were caused by a present parent or sibling (2% dads, 2% siblings, 0% moms).

Families also spent time during the calls talking about the technology. In fact, this occurred in 76% of the calls. In calls that were disrupted by at least one technical problem, the average amount of time spent talking about the technology equaled the average duration of the technical problem itself (43 seconds). Not all calls that had technical problems included this kind of talk, and not all instances of technological talk occurred in calls that had technical problems; however, the two were correlated ($r = 0.50, p = 0.01$).

**Overall Attention**

The baby’s overall attention to the call was assessed, where attention rate was defined as the total duration of the baby’s attention to the call, divided by the total coded duration of the call.
The studied behaviors were tested for their predictive power of the child’s overall attention to the video call. Given the small sample size, only two predictors were chosen for inclusion in a regression. First, because attention for this study was defined as screen-directed behaviors, across-screen JVA was chosen for the regression, as it is the most overt way of directing a child’s attention to the screen itself; within-screen JVA, while it may be related as well, could also be used to direct the child’s attention to a third object in his or her environment. Second, the relative’s sensitivity was chosen because the contingency of adults while interacting with babies has been shown to relate to infant attention (e.g. Hains & Muir, 1996), and our measure of sensitivity includes this contingency as an important indicator. While parental sensitivity may be related as well, the relative’s sensitivity was chosen because their sensitive efforts to engage the child would overtly direct the child’s attention toward the screen (where the relative’s image was located), whereas the parent’s efforts could potentially direct the child’s attention away from the screen and toward the parent instead.

There was only one specific hypothesis regarding the effect of the studied behaviors on the child’s attention: it was predicted that across-screen JVA and the sensitivity of the remote grandparent would both be significant predictors of the child’s attention. To test this hypothesis, a multiple regression was performed with these two variables entered as predictors of the child’s attention rate. The results of the regression model are presented in Table 4.

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>SE</th>
<th>β</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.45</td>
<td>0.33</td>
<td></td>
</tr>
<tr>
<td>Relative’s sensitivity</td>
<td>0.09†</td>
<td>0.05</td>
<td>0.35†</td>
</tr>
<tr>
<td>Across-screen JVA</td>
<td>0.54*</td>
<td>0.24</td>
<td>0.40*</td>
</tr>
<tr>
<td>R²</td>
<td>0.37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>6.32**</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

† marginal; *p ≤ 0.05; **p ≤ 0.01; ***p ≤ 0.001
This regression model was statistically significant \((p = 0.007)\) and accounted for 60% of the variance in the babies’ attention rates. The usage of across-screen JVA was a significant predictor of attention, but the remote grandparent’s sensitivity was only a marginal predictor \((p = 0.06)\), partially confirming the hypothesis.

The relationship between attention and the other studied behaviors – physical contact, technical problems, and JPP – were explored without specific hypotheses and using corrected p-values.\(^{16}\) Correlation analyses were conducted between the baby’s attention rate and the following measures: the parent’s sensitivity, the frequency\(^{17}\) of affection given to or received from the baby, the frequency with which the baby attempted to touch the call device, the families’ level of discouragement toward device touching, the frequency of technical problems, and the degree to which JPP included the baby (versus a sibling) (See Table 5). Only the discouragement of device touching had a significant (positive) relationship with the baby’s attention rate; the degree to which the baby participated in JPP was marginal at traditional levels, but non-significant using the corrected value; the other behaviors were not significant at either the traditional or corrected p-value.

<table>
<thead>
<tr>
<th>Table 5(^{18})</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Correlation coefficients and significance levels for attention</strong></td>
</tr>
<tr>
<td>Parent’s sensitivity</td>
</tr>
<tr>
<td>Frequency of baby affection</td>
</tr>
<tr>
<td>Frequency of baby device-touch attempts</td>
</tr>
<tr>
<td>Discouragement of device touching</td>
</tr>
<tr>
<td>Frequency of technical problems</td>
</tr>
<tr>
<td>JPP with baby</td>
</tr>
</tbody>
</table>

\(^{16}\) Because six parallel tests were conducted, the p-value for significance was adjusted to \(p=0.008\).

\(^{17}\) All frequencies were entered in the number of such behaviors per minute, to control for varying call lengths

\(^{18}\) Only correlations with attention are presented here, to avoid decreasing the adjusted cut-off p-value any further; however a purely descriptive correlation matrix table is presented in Appendix B.
Discussion

This observational study examined how families respond to their babies’ developmental challenges when using video chat with remote grandparents. Despite numerous obstacles, including the screen mediation of JVA, the loss of physical contact between the baby and relative, and contingency delays due to technological problems, families appeared to maintain successful and supportive interactions between the babies and their virtual grandparents.

Findings

Joint visual attention. Attempts at joint visual attention were remarkably successful, notwithstanding the babies’ ages and the unique challenges that video mediation presents. This fact suggests that parents and relatives are modulating JVA attempts well and responding appropriately to their babies’ developmental needs. Surprisingly, this support did not often occur in the form of real-time parental assistance during specific JVA instances. Instead, babies, parents, and remote grandparents appear to simply use the video mediated JVA types that are most accessible and appropriate for the babies’ ages, using within-screen JVA far more often than across-screen JVA.

Furthermore, the babies themselves initiated the more accessible within-screen JVA than the more challenging across-screen JVA, so overt assistance from adults may not have been necessary. Babies initiated more JVA instances as their ages increased, and older babies were more likely to initiate at least one instance of across-screen JVA than the younger babies. These results, combined with the number of children in each age group who initiated at least one instance of within- or across-screen JVA, are suggestive of a possible age-related developmental trend in initiating within- and across-screen JVA: The youngest babies rarely if ever initiate JVA of any kind during the calls, the majority of babies initiated within-screen JVA only, and just the
older babies added initiations of across-screen JVA to their repertoire. This trend is in agreement with the hypothesized difficulty level of these two types of video mediated JVA, based on the existing literature on JVA development in babies.

**Joint pretend play.** Given that joint pretend play is fairly complex developmentally, it was used surprisingly often and successfully in these interactions. Most of it was shared between the child and the remote grandparent. Cross-screen pretend play is a relatively new phenomenon, especially in the informal home environment, and may become more common in children’s lives as video chat continues to become more accessible.

**Physical contact.** While babies and their remote grandparents were unable to make physical contact with one another, parents demonstrated a great deal of physical affection with their children during the interactions. Some experimental studies examining infant and toddler responses to video chat (Tarasuik, Galligan, and Kaufman, 2011; Tarasuik et al., 2013) compare children in a face-to-face interaction condition with the condition of a toddler left alone in a room with a live video feed. It is important to remember, however, that for babies and toddlers, the real-world alternative to face-to-face interaction with a remote relative is a supported video call interaction in the presence of one or more familiar adults. Babies are still receiving plenty of physical contact, even if it does not come directly from the remote relative. In some cases, parents even became creative in acting as a proxy for the physical contact of the grandparent. In one case, for example, while a baby’s grandmother recited the nursery rhyme “this little piggy went to market” on screen, the mother pinched the baby’s toes at the appropriate times.

**Technical problems.** Technical problems did not appear to affect the babies’ overall attention to the calls, despite the fact that they often led to the reduction or absence of contingency on the part of the grandparent. This may have been due to the involvement of
parents in engaging their children during these periods. Parents were observed scaffolding positive responses to delays and other technical problems, and engaging the children’s attention by reframing the problems. For example, in one instance a grandmother’s image froze and then disappeared for 45 seconds, during which the physically present mother modeled a game of hide-and-seek for her 23-month-old child, asking him, “Where’d she go? Where’s Yaya?” The boy imitated her, smiling at the screen and saying, “Where’s Yaya?” in a playful tone. When the grandmother’s image returned to the screen, both the mother and child exclaimed, “There she is!” as though the delay was intended from the start as part of a game. With the help of his mother, it appears that the baby interpreted this incident not as a moment of disrupted contingency but as a moment of highly suspenseful, highly contingent play. By reframing these moments and modeling positive responses, parents may have decreased the losses of attention that would be expected from children who are alone (e.g. in a laboratory environment) under these circumstances.

Attention. It was unclear at the beginning of this study whether children under two could – and would – pay attention to extended video mediated interactions. In the observed video calls, babies attended to the interactions for a little under half their lengths, while moving from place to place at moderate to high levels. The amount of across-screen JVA that was used in calls was a significant predictor of the child’s attention duration, perhaps because parents and grandparents used it as a strategy to redirect the child’s attention to the call. The remote grandparent’s sensitivity to the baby during the call was a marginal predictor of the baby’s attention duration, and the degree to which JPP instances included the baby was also nearly significant. More research systematically examining the role of these variables in attention is needed.
Finally, as the correlations with the child’s attention demonstrate, the degree to which parents verbally discouraged their babies from touching the call device had a highly significant, moderate to strong relationship with the child’s attention. The directionality of this relationship is unclear. As the children in this sample were too young to have developed strong self-regulation skills, parents would have had to rely on their own strategies for preventing the children from touching the screen. According to Rothbart, Posner, and Kieras (2006), caregivers of babies in these early stages of self-regulation development use distraction strategies to reorient the child’s attention away from the prohibited object. Unfortunately, in this case, the prohibited object (the call device) is also one around which the interaction is centered and toward which parents encourage the child’s attention overall. It may be that the parents find it difficult to distract the child from touching the screen while also maintaining their attention to the video call. In other words, perhaps children who are very interested in the call device attempt to touch it more, leading to more touch discouragements; but since the normal parental strategy of redirecting attention elsewhere may not be used in these situations, the interested child may persevere with touch attempts – which, as a screen-directed behavior, would manifest as more attention to the call.

This potential parental conflict (i.e. keeping the child from touching while also maintaining the child’s attention to the video call) may lead to the termination of calls: while the baby’s touch attempts were not significantly negatively correlated with call length ($r = -0.32, p = 0.11$), the magnitude of the negative correlation there, combined with the small sample size, suggests that the relationship should be studied further.
Touch discouragements were also not correlated with call length ($r = -0.09, p = 0.65$), but this could be because the parents relied on discouraging touching via non-verbal methods (e.g. holding the mobile device farther away from the child, see Figure 5), and such non-verbal discouragements were not coded. In other words, we are likely underestimating the degree to which parents discourage such attempts. This makes it difficult to pursue more detailed questions regarding the relationship of discouragements to outcomes like call length.

**Limitations**

The current study had a number of limitations. First, the sample was small and consisted of families from the DC-metro area, so the results should be interpreted with limited generalizability in mind. Future studies should recruit a more representative sample of families from broader, more diverse geographical areas. Second, while every attempt was made by the researchers to remain unobtrusive during the calls, it is still possible that their presence changed the behavior of the participants. The results should be read with this in mind.

Third, there was a limitation in the way that technical problems were coded in this study. In order for coders to reach reliability, only significant technical problems with a clearly defined start and end point could be coded (e.g. the screen goes black for a few seconds and then returns to the normal display). For this reason, more subtle technical difficulties were impossible to code, even if they were nearly continuous. This was most problematic in calls that experienced a
poor connection: in such cases, the image of the screen relative may be obviously jerky or lagging throughout the call, but the problem was not coded. It is quite possible that these delays had an effect on the child’s attention, but there was no way to test this hypothesis with precision in the current study. In the future, new (and unobtrusive) strategies for precisely measuring connection delays in the field should be devised and employed during observational studies.

**Future Directions**

Studies like this one can benefit from the use of growth curve analysis (GCA), which can be used to monitor the variables over time throughout the video calls. GCA can examine these variables at the level of each individual instance of behavior, thus increasing the variation and sample size available for answering research questions. For example, it could be used to answer the following questions: Mothers and their babies tend to initiate the same JVA types; are mothers explicitly modeling JVA types for their babies? Is there an upsurge in across-screen physical affection like screen kisses at the very end of calls, when the families are saying their good-byes? Do the babies’ attention levels wane over time during the calls? And how might events like technical problems, the sharing of physical affection, and the initiation of JVA or JPP affect this time trend? In the future, GCA should be used to explore such questions.

There were also a number of opportunities for future research that were specific to the particular studied behaviors. These are described in the sections below.

**Joint visual attention.** Future studies should consider the relationship between theory of mind (ToM) development and the babies’ ability to participate in successful video mediated JVA. Under certain experimental conditions, it can be demonstrated that babies as young as 18 months can infer what others are able to see (Poulin-Dubois, Sodlan, Metz, Tilden, & Schoeppner, 2007), which is an early ToM skill. Such studies rely on infant looking-time to infer
the expectations of their preverbal participants, even for those up to 24 months of age. In this study, a 23-month-old boy was observed several times pointing a smartphone screen toward objects and toys to show his remote grandmother the objects (see Figure 6). While he sometimes required assistance to do so, his actions demonstrated a burgeoning ability to understand what his remote grandmother was capable of seeing through the built-in camera. Furthermore, this child’s mother frequently used “mind-minded” language to help the child understand what the grandmother was able to see, thus creating an environment that supports the development of ToM (Dunn, Brown, Slomkowski, Tela, & Youngblade, 1991).

Future studies should examine several related questions. First, could a child’s ability to manipulate a camera’s viewpoint (i.e. to show something to a viewer) be used as a novel method of assessing ToM understanding among pre-verbal children? Second, how does understanding what another individual can see relate to understanding what can be seen by that individual through a camera? Are their developmental trajectories similar or distinct? Third, does experience using video chat, particularly with a parent who verbally assists by using mental state language (e.g. “She can’t see it there. Turn her this way so she can see.”), support ToM development? Fourth, does the delivery method of a ToM task affect the outcome of the test? In
at least one of the seminal ToM studies in the field, a false belief task (in which children viewed
an object being hidden) was presented to 3-year-olds via non-contingent video (Moses & Flavell,
1990). The 3-year-olds in this study did poorly on the test – but would these results have been
different if the hiding scenario were demonstrated via live video feed? Research has
demonstrated that toddlers succeed more often in an object retrieval task when they see the
object hidden via live video feed than via non-contingent video (Troseth & DeLoache, 1998),
which suggests that the results of Moses and Flavell’s (1990) study may have been influenced by
their delivery method. A replication of this study that includes a live video feed condition would
be informative. Future studies in this area would mutually benefit research in both ToM and in
children’s media.

**Joint pretend play.** The current study provided a preliminary glimpse at video mediated
pretend play, but more research in this area is needed. In particular, many JPP episodes revolved
around objects of attention, like shared food, so future research investigating the relationship
between JPP and JVA during video calls is necessary.

Furthermore, while many of the pretend play episodes were coded as “successful” (i.e.
the individual invited to play reciprocates appropriately), it remains unclear how much of the
pretense in these cross-screen scenarios is understood as impossible by the children.
Pierroutsakos and Troseth (2003) demonstrated that babies’ under the age of 2, particularly those
under 15 months, struggle to understand the representational nature of screens. It is possible that
the younger children in the sample experienced these “pretend” scenarios differently than the
older children, perhaps believing them to be real. Evidence of this possibility was observed on a
few occasions. For example, one 22-month-old child “fed” raisins to her grandfather, and also to
the grandfather’s dog, throughout the interaction (See Figure 7); however, she also routinely
walked behind the tablet screen, apparently searching for her grandfather on the other side. When her mother attempted to explain to the child that the grandfather was not physically present, she asked her child, “Remember? Where does PopPop live?” The child replied emphatically, saying twice, “Right there,” pointing to the screen.

Figure 7. Sequence: Baby offering raisin through the screen. Left panel is of the child; middle panel is the child offering toward the remote adult; right panel is of the remote adult pretending to eat the raisin.

This is reminiscent of the anecdote referenced in Pierroutsakos and Troseth’s article (2003) describing a 2-year-old getting a real paper towel in response to viewing an egg breaking on television. Further research is needed to systematically explore the development of representational comprehension in these complex and interactive cross-screen JPP episodes.

Physical contact. This study documented the amount of physical affection exchanged between the physically present participants in the video call. In the future, across-screen physical affection can be investigated further by coding the most common types of pretend activities used, to learn how often pretend physical contact between children and their remote grandparents (e.g. screen kisses, passing toys or food through the screen to one another) occurs and its relationship to outcomes like the child’s attention.

Conclusion

Prior to this study, there was no existing research studying natural video chat use by babies under two in their own homes. It was therefore unclear whether babies would be capable
of using these technologies successfully to interact with others. This study intended to bridge this gap by examining how babies and their families responded to the challenge of using video chat at home. Overall, the interactions were surprisingly successful, involving high levels of shared attention, pretend play, and physical affection, despite barriers to their use and the frequent occurrence of technical problems. This is an encouraging result, given the high rates of video chat usage among families with babies and toddlers.
Chapter IV: Infant Emotional Engagement in Face-to-Face and Video Chat Interactions with their Mothers

Introduction

The accessibility and affordability of modern communication media have made it possible today for separated families to maintain contact with one another at a distance. For preverbal children, however, the options are limited. Studies have demonstrated that audio-only communication, like telephones, is not particularly effective for very young children (Tarasuik, Galligan, and Kaufman, 2013) and remains difficult for them to use even through age seven (Ballagas, Kaye, Ames, Go, & Raffle, 2009). Video chat offers an important alternative for families with young children, and especially those who are preverbal, as it offers the real-time interaction of phone interactions, while also providing the visual and non-verbal input of face-to-face communication. Recent evidence suggests that many families use video chat regularly with their babies and toddlers (Chapter II, this dissertation), and that they are able to use it fairly successfully at home (Chapter III, this dissertation). It’s unclear, however, whether babies are as emotionally engaged in these interactions as they are in face-to-face interactions. To examine this systematically, a controlled, experimental study is required.

Mother-infant interaction is known to be emotionally significant in face-to-face contexts (Tronick, Als, Adamson, Wise, & Brazelton, 1978), so it provides a strong standard for comparison to video chat. Sensitive, contingent responses from mothers to their babies are known to play an important role in both babies’ social-emotional development (Stayton & Ainsworth, 1973; Grolnick, McMenamy, & Kurowski, 1999) and in heightening babies’ positive responses during social interactions (Stern, 1993). Furthermore, there is some evidence suggesting that babies under one year of age are emotionally sensitive to very subtle changes in
contingency and reciprocity in their mothers’ behavior (Murray & Trevarthen, 1985; Gusella, Muir, & Tronick, 1988; Nadel, Carchon, & Kervella, 1999).

Prior research comparing infant emotional responses to their mothers on screen and face-to-face is limited, but a few existing studies can inform our expectations. So far, it has been demonstrated that presentations of events on screen do, indeed, have the power to affect babies emotionally, albeit with less intensity that events presented in-person (Diener, Pierrootsakos, Troseth, & Roberts, 2008). Furthermore, a comparison across still-face studies has shown that babies take longer to smile at their mothers when the procedure is presented via live video feed than when in the typical, face-to-face fashion (Gusella et al., 1988). Some older babies have also been shown to smile and attend more to their mothers when presented via live video feed than when presented in a video replay (Bigelow, MacLean, & MacDonald, 1996). Given the scarcity of research and the different methodological approaches across the few existing studies in this area, however, it is difficult to form a coherent idea of how babies respond to loved ones in face-to-face and video mediated scenarios. A more systematic approach is needed.

One study in particular made important steps in this direction, and it is worthwhile to explore it in greater detail. Hains & Muir (1996) conducted two studies comparing the responses of 5-month-olds to individuals in video mediated conditions. First, they examined how babies reacted to their mothers when they interacted as usual via contingent (i.e. live video feed) or non-contingent (i.e. pre-recorded) video. They found that there were no differences in the babies’ visual attention or number of smiles towards their mothers in these screen conditions. This study revealed that babies may be tolerant of short periods of non-contingency from their mothers, whose interaction style is very familiar to them, so long as the mother continues to use her normal interactive style. This study, however, like others of this kind (Murray & Trevarthen,
1985; Nadel, Carchon, & Kervella, 1999), did not include a presentation of the mother in a face-to-face condition, so the results can only be interpreted for mothers presented on screens. For this reason, it is still unclear how video mediated maternal sensitivity and responsiveness would compare to its well-studied face-to-face counterpart.

Hains and Muir (1996) conducted a second study, this time comparing babies’ responses to presentations of a stranger, now in all three conditions (i.e. contingent video, non-contingent video, and face-to-face). Here, the infants were more visually attentive to both of the contingent interactions with the stranger – face-to-face and contingent video – than to the video replay version. This suggests that infants as young as 5 months have the ability to detect very subtle differences in contingency between these interactions. Additionally, while these infants had an equal level of visual attention to both the face-to-face and the live video link interactions, infants smiled earlier (as in Gusella et al., 1988) and more often in the face-to-face interaction, indicating a more positive response to individuals who are presented in-person than on screen. While Hains and Muir’s (1996) second study has the advantage of comparing all three conditions side-by-side, it unfortunately uses a stranger for the interaction.

The present study aims to extend the Hains and Muir (1996) studies in four ways. First, Hains and Muir (1996) used strangers instead of mothers in their study comparing all three conditions. In the present study mothers will be used instead, but retaining the strong approach of comparing all three conditions side-by-side.

Second, Hains and Muir (1996) allowed mothers to interact as usual with their babies, which likely presented a great deal of variation in the interactions. Following Diener et al. (2008), the present study will use a more precisely defined peek-a-boo interaction procedure for a stronger comparison across groups. Peek-a-boo was chosen for several reasons. It is infants’
growing understanding of contingency in parent-child interactions that leads to their pleasure in interactive games like peek-a-boo (Roggman, 1991). Furthermore, it is developmentally appropriate for babies under 1 year of age. By around 5.5 or 6 months of age, infants have the working memory ability to be able to participate accurately in a peek-a-boo game (e.g. they anticipate correctly that the face will show up again in the same location as before) and it is engaging for them (Reznick, Morrow, Goldman, & Snyder, 2004); and by 6-8 months, infants not only have specific expectations about peek-a-boo and enjoy it, they also smile less when it doesn't meet those expectations (Parrott & Gleitman, 1989). Finally, peek-a-boo has been shown to remain popular among babies across the 6 to 12 month age range, and mothers have been shown to effectively modulate it in a way that is developmentally appropriate and engaging as their infants become more responsive across those months (Gustafson, Green, & West, 1979). This makes it a strong choice for this study, in which mothers, rather than a trained experimenter, will be used in the presentation of the experimental stimulus.

Third, previous studies in this area, including Hains and Muir (1996), have all used a repeated measures design, using a live video – video replay – live video order to test their hypotheses. Repeated measured designs are very strong in general; however, in these cases they have also led to results influenced by order effects and the babies’ waning interest in the stimulus. This study will incorporate a between-subjects design to help parse out these issues.

Fourth, prior studies, including Hains and Muir (1996), have relied exclusively on expressive behavior as a measure of emotional reactivity; however reactivity is not limited simply to expression. Particularly among samples that are unable to report their emotional experiences verbally, measures of physiological reactivity provide an additional measure of emotion. In the present study, measures of skin conductance levels, heart rate, and respiration
rate will be collected, as these are well-established indicators of emotional reactivity (Boiten, Frijda, & Wientjes, 1994; Levenson, 2003). Furthermore, both skin conductance and cardiac measures are used regularly to assess adult reactivity to digital media presentations (Ravaja, 2009), and their use with infants has been previously established (Hernes, Morkrid, Fremming, Odegarden, Martinsen, & Storm, 2002; Ham & Tronick, 2008; Lansink & Richards, 1997).

An experimental study was conducted with babies between the ages of 6 and 12 months to systematically compare the infants’ emotional responses to their mothers as presented face-to-face, via video chat, and via non-contingent video. Combining the results of the three previous studies, the following hypotheses can be made:

**Hypothesis 1:** If the results of Hains and Muir (1996, study 1) are replicated, there will be no difference in babies’ emotional reactivity to their mothers in the two screen conditions, as reflected in equal amounts of smiling and similar levels of physiological responsiveness across the groups; there will also be equal levels of attention across groups. The alternative will be tested: namely, that the use of a highly engaging peek-a-boo procedure will result in babies paying more attention to their mothers and smiling more face-to-face and via video chat than they do via non-contingent video (Bigelow et al., 1996; Hains & Muir, 1996, study 2); the physiological responsiveness will reflect this heightened emotional reactivity.

**Hypothesis 2:** As in Gusella et al. (1998), infants will smile earlier during the face-to-face interaction with their mothers than in the screen conditions.

**Hypothesis 3:** Given the strong relationship between maternal sensitivity and infant social-emotional responsiveness, it is expected that the mother’s sensitivity will be an
important predictor of the baby’s emotional reactivity, as reflected in the number of baby
smiles and in physiological responsiveness.

Method

Participants

Forty-nine infants between the ages of 6 and 12 months, their mothers, and a familiar
second caregiver were recruited for a “peek-a-boo study” from the D.C.-metro area via flyers,
listservs, community events, and word of mouth. The second caregivers were fathers (82%),
grandmothers (10%), nannies (4%), a close friend (2%), and a mother (2%). The babies’ mean
age was 8.9 months (SD = 1.86), 45% were female, and 61% had previous experience using
video chat (defined as having used it once a month or more). Fifteen additional infants (23%)
were tested but were not included in the final sample due to crying within the first minute of the
interaction period (n = 9), experimenter error or equipment malfunction (n = 4), and parent error
(n = 2). It is typical for studies of this kind to have attrition rates above 20% (Lansink &
Richards, 1997).

Apparatus

**Baby interaction booth.** An interaction booth for the study was designed following
Diener et al. (2008). Babies faced a black, foam-board display booth, set on a table at roughly
eye level and 4 feet away from the baby (Figure 8). The booth had a 10” x 16.5” opening in it, in
which both the face-to-face and video interactions were presented: For the video interactions, a
19-inch, black, flat-screen monitor was placed just behind the opening; for the face-to-face
interactions, mothers sat on a 9-inch tall stool, placed such that her face was 30 inches away
from the opening. Above the opening was a small rectangular hole, in which a black web camera
was placed to record the babies’ behavioral responses and transmit them to mothers in the video
chat condition. On either side of the opening, black speakers were placed behind a small rectangular hole, to allow the mothers’ audio feed to be heard by the babies in both screen conditions. On the opposite side of the booth from the baby, a black poster curtain was used by an experimenter to cover the opening prior to the interaction period; this curtain was lifted to signal the beginning of the interaction period, and then dropped again to signal the end of the interaction. To prevent distractions, white fabric curtains were hung on either side of the booth, covering from floor to ceiling and from the edges of the booth to the walls of the lab space.

![Image](image1.png)

*Figure 8. Baby interaction booth. Top left: Father and baby sit facing the interaction booth with curtain down. Top right: Baby’s point of view, facing the interaction booth with curtain down. Bottom left: View of a model playing peek-a-boo in the face-to-face condition (reenactment). Bottom right: View of a model playing peek-a-boo in the screen conditions (reenactment).*

*Mother interaction booth.* A booth was also created for mothers in the two screen conditions, which was placed in an adjacent room (Figure 9). Mothers faced a black, foam-board booth, which was lined with sound-absorbing foam material to prevent her voice from being audible in real-time to her baby in the next room. Within the booth was a 19-inch, black, flat screen monitor, upon which a black webcam was placed to record the mothers’ behaviors and
transmit them to the baby. The mother sat in a chair located such that the camera was 30 inches away from her face, a position that produced an image of the mother than was equal in size, from the babies’ perspective, to the mothers’ faces in the face-to-face condition. A white noise machine was placed near the shared wall between the adjacent rooms, which was activated for the duration of the experiment to prevent the mother and baby from hearing one another through the wall.

Figure 9. Mother interaction booth. Left: mother plays peek-a-boo in the video chat condition. Right: Mother prepares to play peek-a-boo in the video condition, facing a photograph of her own smiling baby.

Measures

Physiological measures. A system consisting of a computer, Acqknowledge 3.9.1 software (Biopac Systems, 2009), and bioamplifiers (Biopac Systems) were used to obtain continuous recordings of the babies’ physiological reactivity.

To measure heart rate, electrodes were placed in a bipolar configuration on opposite sides of the chest. Heart rate was measured in beats per minute. To measure skin conductance level, a procedure adapted from Ham and Tronick’s (2008) method for collecting skin conductance in infants was used. A constant-voltage device passed a small voltage between electrodes attached to the plantar surface of the foot (on the heel) and on the infant’s thigh. Skin conductance level was measured in microsiemens. To measure respiratory rate, a respiration belt was placed around
the upper torso. Respiratory rate was measured in respirations per minute. Acknowledge software extracted raw data and produced wave-form transformations, peak detection, and graphic display for each channel. Higher levels of heart rate and skin conductance levels indicate higher levels of physiological reactivity.

A trained research assistant removed artifacts from the data and calculated averages for the baseline and interaction period of the experiment. The average during baseline was then subtracted from the average during stimulus to create change scores, and all analyses of the physiological data were conducted using these scores. Due to equipment malfunction during the collection of skin conductance activity, the sample size for skin conductance analyses was smaller than that of the other analyses considered ($n = 36$).

**Infant Temperament Questionnaire.** Infant temperament was assessed using the *Infant Behavior Questionnaire – Revised, Very Short Form* (Putnam, Helbig, Gartstein, Rothbard, & Leerkes, 2014), which is a parent-report measure designed for infants between the ages of 3 and 12 months of age. The 37-item form measures three components of infant temperament – Positive Affect, Negative Affect, and Orienting – using questions about the baby’s behavior during the past week. The questionnaire took mothers approximately 5 minutes to complete.

**Design**

Mothers were randomly assigned to play peek-a-boo with their babies in one of three conditions, allowing for age, sex, and video chat experience to be roughly evenly distributed across groups. The following three conditions were used:

**Face-to-face:** 30 seconds prior to the beginning of the interaction period, the mother was silently escorted into the room where the baby was located, on the opposite side of the curtain and booth from the baby so she could not be seen. A video camera on a tripod
behind the baby’s right shoulder recorded the mother’s face during the interaction, and a webcam in the booth setup recorded the baby’s face.

**Video chat:** The mother sat at an interaction booth in an adjacent room, facing a computer monitor with a webcam above it. The use of nonintrusive ear buds allowed her to hear the baby. The webcams in each booth were used to both transmit and record the images of the mother and baby to one another. Prior to the interaction period, a black curtain obstructed both the baby’s and mother’s views of one another. This curtain was raised as a signal to the mothers to begin interacting.

**Video (non-contingent):** The mother sat at an interaction booth in an adjacent room, facing a computer monitor with a webcam above it. Mothers wore nonintrusive ear buds in this condition to maintain uniformity with the video chat condition; however the ear buds only played white noise. Webcams recorded both the mother’s and baby’s faces, while also transmitting the mother’s image to the baby’s monitor. The mother’s computer monitor was turned off for the duration of the experiment; however, a printed 4” x 6” photograph of her smiling baby, acquired prior to the study, was placed on the monitor at the same location where the baby’s face would appear on the screen in the live video feed condition. Mothers were instructed to play peek-a-boo with their child while looking at and engaging with the photo before them. The mother held the end of a piece of yarn, which was connected to the setup in the adjacent room, in her left hand while waiting to begin; she was signaled to start interacting by a tug on the yarn.

**Procedure**

Parents were sent one of three instructional videos, based on their condition, prior to the study. Upon arrival at the facility, mothers and babies were separated during a brief distraction
period to avoid infant anxiety. Mothers and babies were escorted separately into two adjacent rooms for the warm-up and instruction periods.

**Baby procedure.** A researcher spent up to 10 minutes warming up with the baby by playing with toys and books on the floor of the lab space. During this time, the researcher also obtained informed consent from the second caregiver. After the warm-up period, the baby was seated on the lap of the caregiver for the duration of the experiment. The researcher then positioned the respiration belt and placed the cardiac and skin conductance electrodes appropriately. The procedure typically took less than 5 minutes.

The baby and caregiver then sat quietly for 90 seconds to establish the baby’s baseline physiological responsiveness. Because babies were unable to stay still and content for this period without any stimulation at all, babies were provided with a small, relatively non-stimulating toy (a block, a cup, or a small book) to hold during the baseline period. The caregiver was instructed to stay quiet and still throughout the baseline period, only interacting with the baby (using a gentle method like mild rocking or humming) if the baby began to fuss.

After the baseline period, the curtain was lifted by the researcher, and babies were presented with their playing mothers for 2.5 minutes; this period was truncated if the baby began fussing continuously ($n = 5$). The caregiver was instructed to remain still and quiet during the interaction period, reacting neither to the mother’s presentation nor to the baby’s responses.

**Mother procedure.** During the warm-up period, informed consent was obtained from the mother and she was given instructions in an adjacent room. The mother was instructed to begin playing peek-a-boo with her baby immediately when the curtain was lifted (or when the yarn was tugged, see section *Design*) in whatever way she normally would at home. A black handkerchief

---

19 The minimum interaction time was one minute. Babies that began crying prior to one minute were dropped from analysis.
was provided as a prop, but the mother was instructed to use it only if she and her baby would normally use props like these at home.

At the conclusion of the interaction period, the mothers and babies were reunited for a short play period, during which the mothers completed the infant temperament questionnaire and the families were fully debriefed about the purpose of the study. Babies received a small toy and certificate to take home, and families were compensated $15 for their time and travel.

**Behavioral Coding**

Two coders were trained to identify behaviors of interest for the study. The coders were not informed of the hypotheses of the study; however, given that the videos of mothers in the face-to-face condition were, of necessity, recorded at a different angle (i.e. from behind the baby’s shoulder rather than from directly head-on, as with the webcam), experimental differences could not be completely hidden.

Both coders coded 22% of the recordings for reliability purposes, and an overall inter-rater reliability of ICC = 0.95 or higher was attained. Any behavior that did not appear in at least fourteen sessions was dropped from all further analyses. After reliability was reached, each coder was responsible for coding 50% of the recorded sessions. The total frequency or duration of each behavior per call was calculated.

The code definitions and their corresponding reliability coefficients are presented below.

**Baby: Number of smiles.**<sup>20</sup> Smiles were defined as when the corners of the baby’s mouth were pulled upward, not just laterally (as distinguished by Ekman & Friesen, 1978). Interrater reliability for this code reached ICC = 0.95.

---

<sup>20</sup> The frequency of baby smiles and maximum negative affect were significantly correlated \( r = -0.31, p = 0.03 \); however, smile frequency and latency to smile were not \( r = -0.27, p = 0.07 \).
**Baby: Latency to smile.** The duration between the start of the interaction period (signaled by the lifting curtain) and the first infant smile. Inter-rater reliability for this code reached ICC = 0.99.

**Baby: Maximum negative affect.** The maximum intensity of the baby’s negative vocalizations was assessed once for the baseline period and once for the interaction period using Braungart-Ricker & Stifter’s (1996) 5-point scale: 0 (no negative vocalizations); 4 (shrieking, hysterical crying). Inter-rater reliability for this code reached ICC = 0.99.

**Baby: Attention.** The baby’s attention was defined as whether the baby’s gaze was directed at the mother/screen or not (eyes on / eyes off), as assessed using the webcam directly above the opening in the booth. The total duration of “eyes on” was summed to create the total duration of baby attention. The number of times the baby’s eyes turned away and then returned back to the mother or the screen, an index called “distractibility” in the analyses to follow. Inter-rater reliability for this code reached ICC = 0.98.

**Mother: Intentional facial occlusions.** The incidence and duration of events in which the mother intentionally hid her face for playing peek-a-boo was recorded. Occlusion was defined as when the mother put her hands together in front of her face, used the handkerchief to obstruct her face, intentionally moved beyond the baby’s view to hide (usually laterally), or otherwise deliberately hid her face for the game. Inter-rater reliability for the incidence of occlusions reached ICC = 0.99. Inter-rater reliability for the duration of occlusions reached ICC = 1.00.

**Mother: Unintentional facial occlusions.** In the two screen conditions, mothers occasionally disappeared from view unintentionally (Figure 10) by leaning outside the camera view (usually by leaning forward, below the camera). The unintentional nature of these
occlusions was often signaled by their inappropriate timing (e.g. the mother pops out from behind her hands [an intentional occlusion], but leans so far forward upon popping out that she moves below the camera view). Both the incidence and duration of these events were recorded. Inter-rater reliability for incidence reached ICC = 1.00. Inter-rater reliability for duration reached ICC = 1.00.

![Figure 10. Unintentional facial occlusion.](image)

**Mother: Sensitivity.** The mother’s sensitive behavioral style during the interaction period was assessed using the *Parental Sensitivity Scale* from *The Emotional Availability Scales, abridged Infancy/Early Childhood Version* (Biringen et al., 2000). Coders provided a single score (between 1 and 9) for each mother. It is important to note that “sensitivity” on this scale does not measure exact, moment-to-moment, contingent responsiveness to the baby. Instead, “sensitivity” here “is highly global and emphasizes behavioral style rather than discrete behaviors” (p. 257). Using this global measure is important for this study because mothers in the non-contingent video condition were incapable of responding to their babies with perfectly timed or perfectly appropriate *discrete behaviors*, due to their inability to see or hear their children. They were, however, able to demonstrate a globally sensitive behavioral style, which includes factors like affect, authentic interest, and playfulness. Inter-rater reliability for this code reached ICC = 0.99.
Results

First, the characteristics of the babies in each condition were assessed for systematic, non-experimental differences (Table 6). There were no differences in age, gender, video chat experience, maternal sensitivity, infant temperament, fussiness, or interaction duration across the three conditions. These characteristics will not be used as covariates in the analyses to follow.

Table 6
Tests for non-experimental group differences

<table>
<thead>
<tr>
<th></th>
<th>Video Chat n = 15</th>
<th>Video n = 16</th>
<th>Face-to-Face n = 18</th>
<th>Test Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>$M = 9.3$</td>
<td>$M = 8.9$</td>
<td>$M = 8.6$</td>
<td>$F(2,47) = 0.60$</td>
</tr>
<tr>
<td></td>
<td>$SD = 1.5$</td>
<td>$SD = 1.8$</td>
<td>$SD = 2.2$</td>
<td></td>
</tr>
<tr>
<td>Gender (female)</td>
<td>47%</td>
<td>56%</td>
<td>33%</td>
<td>$X^2(2) = 1.83$</td>
</tr>
<tr>
<td>Video Chat Experience (yes)</td>
<td>60%</td>
<td>56%</td>
<td>66%</td>
<td>$X^2(2) = 0.40$</td>
</tr>
<tr>
<td>Mother Sensitivity</td>
<td>$M = 8.1$</td>
<td>$M = 7.6$</td>
<td>$M = 7.9$</td>
<td>$F(2,46) = 0.65$</td>
</tr>
<tr>
<td></td>
<td>$SD = 1.2$</td>
<td>$SD = 1.0$</td>
<td>$SD = 1.2$</td>
<td></td>
</tr>
<tr>
<td>Temperament</td>
<td></td>
<td></td>
<td></td>
<td>$F(6,84) = 0.48$</td>
</tr>
<tr>
<td>Positive Affect</td>
<td>5.1</td>
<td>4.8</td>
<td>5.2</td>
<td>$F(2,46) = 1.19$</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>0.6</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>Negative Affect</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>$F(2,46) = 0.00$</td>
</tr>
<tr>
<td></td>
<td>1.1</td>
<td>1.0</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>Orienting</td>
<td>5.0</td>
<td>4.9</td>
<td>5.0</td>
<td>$F(2,46) = 0.08$</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>0.6</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>Number of Fussers</td>
<td>27%</td>
<td>6%</td>
<td>28%</td>
<td>$X^2(2) = 2.94$</td>
</tr>
<tr>
<td>Interaction Length</td>
<td>$M = 156.4$</td>
<td>$M = 153.6$</td>
<td>$M = 158.2$</td>
<td>$F(2,46) = 0.13$</td>
</tr>
<tr>
<td></td>
<td>$SD = 22.6$</td>
<td>$SD = 30.1$</td>
<td>$SD = 26.9$</td>
<td></td>
</tr>
<tr>
<td>Transmission Delay (yes)</td>
<td>33%</td>
<td>50%</td>
<td>--</td>
<td>$X^2(1) = 0.88$</td>
</tr>
<tr>
<td>Unintentional Occlusions</td>
<td></td>
<td></td>
<td></td>
<td>$t(21.1) = 2.04^+$</td>
</tr>
<tr>
<td>Incidence</td>
<td>7.5</td>
<td>1.4</td>
<td>--</td>
<td>$t(18.1) = 2.08^{***}$</td>
</tr>
<tr>
<td></td>
<td>10.3</td>
<td>5.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration (seconds)</td>
<td>15.8</td>
<td>2.4</td>
<td>9.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>23.3</td>
<td>1.4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$marginal; ^* p \leq 0.05; ^{**} p \leq 0.01; ^{***} p \leq 0.001$

The screen conditions included two characteristics unique to their setup: unintentional facial occlusions, and the occasional delay in the transmission of the audio or visual input (<5 ms) during the interaction. There were no differences in the number of sessions that had a transmission delay between the screen conditions. There was no difference in unintentional
occlusions between the two screen conditions (incidence: $p = 0.05$; duration: $p = 0.05$). Due to the marginal nature of this test, however, all applicable analyses were conducted both with and without the duration of unintentional occlusions as a covariate. This covariate did not affect the outcomes of any of the tests, so all analyses are reported without the covariate included.

Next, we conducted a one-factor repeated measures (Experimental Condition X Period [Baseline, Interaction]) MANOVA on the three physiological variables. Analyses revealed a significant effect of Period, $F(3,29) = 5.58, p < .01$. This effect was driven by a significant decrease from baseline to interaction period for heart rate, $F(1,40) = 7.17, p = 0.01$, and a significant increase from baseline to interaction period for skin conductance, $F(1,32) = 6.73, p = 0.01$. There was no difference in respiratory rate between baseline and interaction period, $F(1,40) = 0.25, p = 0.62$.

Finally, the experimental hypotheses were tested. Preliminary ANOVAs investigating the babies’ age and video chat experience

21 Video chat experience was originally measured on a 6-point scale assessing frequency of use (ranging from “Never” to “At least once a week”), but for purposes of statistical power was collapsed here into a dichotomous yes/no variable, with experience being defined as using video chat at least once per month, and no experience as using video chat a few times per year or less.
more face-to-face and via video chat than they do via non-contingent video (Bigelow et al., 1996; Hains & Muir, 1996, study 2); and that the physiological responsiveness would reflect this heightened emotional reactivity.

A one-way ANOVA was used to examine frequency of smiling across the three conditions and revealed that there was no difference in the number of infant smiles across the video chat ($M = 4.86, SD = 4.45$), video ($M = 7.53, SD = 6.36$), and face-to-face ($M = 7.18, SD = 5.93$) conditions, $F(2,43) = 0.95, p = 0.40, \eta^2 = 0.04$. A one-way ANOVA was the used to examine attention levels across conditions. It also revealed that there was no difference in the duration of attention (in seconds) across the video chat ($M = 136.32, SD = 23.92$), video ($M = 140.01, SD = 28.51$), and face-to-face ($M = 132.41, SD = 33.11$) conditions, $F(2,46) = 0.29, p = 0.75, \eta^2 = 0.01$. There was also no difference in the distractibility of the babies across the video chat ($M = 7.80, SD = 6.07$), video ($M = 6.06, SD = 4.43$), and face-to-face ($M = 9.89, SD = 5.27$) conditions, $F(2,46) = 2.24, p = 0.12$.

Three one-way ANOVAs revealed that there were no significant differences in the babies’ change scores for heart rate, respiration rate, and skin conductance (Table 7) across the three conditions.

The null was not rejected for Hypothesis 1, which is consistent with previous findings.

Table 7

| Tests of change score differences in physiological reactivity, by condition |
|-----------------------------|------------------|------------------|------------------|------------------|
|                             | Video Chat | Video | Face-to-Face | Test Statistic |
| Heart Rate                  | M   | SD   | M   | SD   | M   | SD   | $F(2,41) = 0.47, \eta^2 = 0.02$ |
| Respiration Rate            | -1.83 | 7.92 | -4.55 | 7.22 | -3.74 | 7.85 | $F(2,41) = 0.98, \eta^2 = 0.05$ |
| Skin Conductance$^{22}$     | 0.06 | 1.35 | 0.66 | 1.16 | 0.68 | 0.45 | $F(2,33) = 1.38, \eta^2 = 0.08$ |

$marginal; \ast p \leq 0.05; \ast \ast p \leq 0.01; \ast \ast \ast p \leq 0.001$

$^{22}$ Note that the statistical power was reduced, due to its smaller sample size ($n = 36$), for the test of group differences in skin conductance change scores. See section *Preparation of Physiological Data* above.
Hypothesis 2: Latency to Smile

The second hypothesis was that infants would smile earlier during the face-to-face interaction with their mothers than in the two screen conditions (Gusella et al., 1988; Hains & Muir, 1996, study 2). An ANOVA revealed a marginal difference in the latency to smile (in seconds) between the video chat ($M = 12.78, SD = 7.53$), video ($M = 8.40, SD = 3.06$), and face-to-face ($M = 6.97, SD = 8.00$) conditions, $F(2,43) = 3.10, p = 0.06, \eta^2 = 0.13$. While the difference was only marginal, the effect size was large, so post-hoc tests were used to examine whether the effect was in the predicted direction. Post-hoc Bonferroni-corrected tests revealed that the babies in the video chat condition took marginally longer to smile than babies in the face-to-face group ($p = 0.06$), but that there was no difference in the latency to smile between the babies in the video condition and either the video chat ($p = 0.25$) or face-to-face ($p = 1.00$) conditions.

Hypothesis 3: Maternal Sensitivity

The third hypothesis was that the mother’s sensitive behavioral style with the infant would play a role in the baby’s emotional reactivity, over and above experimental condition, as reflected in the number of baby smiles and in physiological responsiveness. It is important to recall when considering the following results that “sensitivity” here does not exclusively refer to the immediacy and appropriateness of mothers’ responses to the babies’ moment-to-moment social/emotional cues; otherwise, the construct would be impossible to apply to the non-contingent video group, in which mothers did not have access to their babies’ visual or auditory cues. Instead, this scale is more global and not based exclusively on the mothers’ responsiveness; instead it is based on the overall behavioral style of the mothers (Biringen et al., 2000).

To test this hypothesis, hierarchical linear regressions were performed using frequency of
smiles, latency to smile, and each of the three physiological measures as dependent variables. In each analysis, experimental condition was entered for Step 1 using two dummy variables, and maternal sensitivity was entered in Step 2. Initial analyses also examined the effect of the interaction terms between experimental condition and maternal sensitivity, but this additional step did not result in significant increases in the proportion of explained variance for the models (all $\Delta R^2 < .02$, ns).

The regression using frequency of smiles was significant ($F(3,42) = 4.94, p = 0.01$) with an $R^2$ of 0.21, and revealed a statistically significant relationship between maternal sensitivity and the frequency of infant smiles, $B = 2.40, SE B = 0.68$, $\beta = 0.47, t = 3.52, p < 0.01$, such that each additional point of rated sensitivity predicted an additional two smiles from the baby during the interaction period. It failed to reveal a relationship between the frequency of smiles and experimental condition.

The addition of Step 2 (maternal sensitivity) to the regression using latency to smile did not result in significant increases in the proportion of explained variance for the model. Step 1 of the model, with an $R^2$ of 0.13, was statistically identical to the one-way ANOVA testing Hypothesis 2 (see above).

None of the steps resulted in significant increases in $R^2$ for the physiological measures of heart rate, respiration rate, or skin conductance.

**Discussion**

Today many families are using video chat with their babies in an attempt to develop and maintain relationships between their babies and a remote relative (Chapter II, *this dissertation*). While families are capable of supporting such calls fairly successfully at home, at least among infants up to 24 months of age and with remote grandparents (Chapter III, *this dissertation*), it
has remained unclear whether babies emotionally engage in these calls with their loved ones in a way that resembles face-to-face interaction. To my knowledge, this is the first study to address this question systematically in a single experiment. It is also the first to employ physiological methods in the examination.

**General Findings**

Initial tests of the physiology measures demonstrated that the peek-a-boo interaction was an effective manipulation of the babies’ emotional and attentional reactivity across the three conditions. Specifically, babies’ heart rates decreased and their skin conductance levels increased between baseline and the interaction periods. A sustained lower heart rate has been shown to indicate sustained attention in babies between 6 and 12 months (Lansink & Richards, 1997), and increased skin conductance indicates greater arousal (Hernes et al., 2002). In other words, the babies appear to have found the interaction period engaging, regardless of their experimental condition.

This study was conducted with three hypotheses in mind. The first hypothesis test yielded results that are consistent with Hains and Muir’s (1996) study with mothers: Babies paid the same amount of attention and smiled just as often when their mothers played peek-a-boo with them face-to-face, via video chat, and via non-contingent video. These equal levels of emotional reactivity were also reflected in similar levels of physiological responsiveness across the conditions. The second hypothesis was partially confirmed: Babies took marginally longer to smile when using video chat than when interacting face-to-face; this difference was statistically significant only when maternal sensitivity was also included in the test. Furthermore, the video condition unexpectedly fell between the video chat and face-to-face conditions on this dimension, differing from neither of them. The third hypothesis was confirmed: Maternal
sensitivity did indeed play a significant role in predicting infant smiles (but not latency to smile), over and above experimental condition.

The results of the first hypothesis, which were consistent regarding the babies’ responses to their mothers on screen, produced one novel result: there were no differences in smiling or attention between the face-to-face and screen conditions. It is important to keep in mind, however, that the face-to-face condition produced a relational situation with limited ecological validity for babies and mothers. The babies sat nearly 5 feet away from their mothers, who were separated from them by a large booth, which made physical contact between the two impossible. Given that mothers and their babies make frequent physical contact during ordinary interactions (Ferber, Feldman, & Makhoul, 2008), the experimental setup itself may have led babies to smile less often than would be normal in true face-to-face scenarios, thus diluting any potential differences between this condition and the others. On the other hand, while every attempt was made by the researchers to remain unobtrusive during the experiment, it is possible that experimenter demand may have changed the behaviors of the mothers in the face-to-face condition in the opposite direction. The experimenter was present in the room with the babies across all the conditions because it was necessary to lift the black curtain at the appropriate time; however, this also meant that the experimenter was also in the room with mothers only in the face-to-face condition. This was not the case for the mothers in the screen conditions because they sat in an adjacent room and participated remotely. It is conceivable that mothers in the face-to-face condition may have behaved more positively – for example, playing peek-a-boo with higher levels of animation – due to the experimenter’s presence. If this is the case, however, the results of this study are all the more striking.
The mixed results regarding latency to smile (hypothesis 2) may be explainable. First, the difference between the video chat and face-to-face conditions was only marginal, which was unexpected. Given the small sample size, the fact that this effect has been demonstrated with statistical significance in at least one other study (Gusella et al, 1988), and the large effect size ($\eta^2 = 0.13$), however, there is reason to believe that the marginal effect has meaning.

Second, the babies’ latency to smile in the video condition fell unexpectedly between those in the video chat and face-to-face conditions, differing from neither of them. Three possible explanations for this were tested. While there were no statistically significant differences in transmission delay or unintentional hides between the screen conditions, it is still possible that these factors still may have played a small role in the test’s outcome. The hypothesis that transmission delay played a role in the babies’ latency to smile was tested. A one-way ANCOVA testing for differences in latency to smile across the two screen conditions, controlling for the presence of a transmission delay (yes/no), was not statistically significant ($F(2,26) = 2.12, p = 0.14$). Then the hypothesis that the presence of unintentional hides played a role in the babies’ latency to smile was tested. A one-way ANCOVA testing for differences in latency to smile across the two screen conditions, controlling for the presence of unintentional hides (yes/no), was also not statistically significant ($F(2,26) = 2.12, p = 0.14$).

Finally, the hypothesis that the mother’s latency to hide for the first time may have affected the babies’ latency to smile across the conditions was tested. A one-way ANOVA revealed that a difference does indeed exist across conditions in how long it took for moms to hide for the first time while playing peek-a-boo, $F(2,46) = 7.84, p < 0.01$. The moms in the video chat condition took the longest ($M = 9.07$ seconds, $SD = 4.50$), followed by moms in the

---

23 The equivalence of the test statistics for these tests is due to rounding and chance. A Chi-Square analysis revealed that the presence of a delay and the presence of unintentional hides were unrelated ($X^2 (1) = 0.39, p = 0.53$).
video condition \((M = 5.06, SD = 3.24)\), followed by moms in the face-to-face condition \((M = 4.17, SD = 3.37)\). The significant differences were between video chat and video \((p = 0.01)\), and video chat and face-to-face \((p < 0.01)\). There was no difference between video and face-to-face. However this difference in the mothers’ “latency to hide” was unrelated to the results for the babies’ latency to smile \((r = 0.16, p = 0.29)\), and it did not add significantly to the \(R^2\) when used as Step 1 in a step-wise hierarchical regression (Step 2: condition).\(^{24}\)

These three testable hypotheses did not explain the unexpected placement of the video condition between the video chat and face-to-face conditions in the babies’ latency to smile. Given the lack of previous research comparing all three conditions simultaneously, any remaining potential explanations must be purely speculative; however, a few possibilities can be offered. One must first take note: Since the marginal difference in babies’ latency to smile was between the video chat and face-to-face conditions, contingency itself was (contrary to expectation) not the factor related to this difference. Instead, the factor that differed between these conditions was whether or not the mother appeared on a screen. If we consider face-to-face interaction as our baseline level of responsiveness, children took longer to smile when interacting via video chat (relative to face-to-face) but not when they viewed their mothers via non-contingent video. The original hypothesis was based on the idea that the video chat condition was more similar to the face-to-face condition in terms of its contingency, and was thus likely to elicit similar reactions from children. Instead, however, it may be that the babies responded based on which screen medium they were more familiar with. Children under two spend nearly 1 hour per day watching television and videos (Rideout, 2013), which means they spend more time being exposed to videos than to video chat (Chapter II, *this dissertation*). It may be that the infants in the video condition did not find their mothers’ non-contingent behavior on screen to be

\(^{24}\) It was also unrelated to the baby’s attention \((r = -0.07, p = 0.63)\) and the mother’s sensitivity \((r = 0.22, p = 0.12)\).
abnormal because their experience with video and television screens supports this expectation of non-contingent screen behaviors (Troseth, 2010). Instead they would have responded to their mothers’ screen contingency as the abnormal behavior – not due to its contingency, per se, but because of their lack of familiarity with social contingency on screens. The additional processing time of the unfamiliar medium may have delayed their smiles during the presentation. This might explain why babies in the video condition fell between the other two groups: their latency to smile may reflect their familiarity with the interaction medium (or lack thereof), with face-to-face being most familiar, video screens being the next most familiar, and video chat being the least familiar.

Hains and Muir’s (1996) study using mothers similarly produced no differences in attention or smiling between the video chat and video conditions. They argued that babies were tolerant of short periods of minor non-contingency from their mothers, given their familiarity with their mothers’ interaction style. They did not include a face-to-face condition, however, so it is unclear whether the responses to non-contingent video in their study would be positioned similarly to these under the same conditions. It is possible, however, that babies are indeed fairly tolerant of short-term non-contingency from familiar loved ones, which would explain the lack of difference between the video condition and the two contingent conditions. This is consistent with naturalistic observations of babies’ and toddlers’ attention levels during video chat interactions at home, which were unaffected by technical problems that decreased the contingency of the interaction (Chapter III, this dissertation).

Finally, while the significant difference across conditions in the mothers’ latency to hide was unrelated to the babies’ latency to smile, exploration of this group difference revealed one additional result, which is of importance for Hypothesis 1. The difference in the mothers’ latency
to hide was related to the babies’ smile frequency. Hierarchical regression was employed for this test, with smile frequency as the dependent variable: The mother’s latency to hide was added in Step 1, and experimental condition was added in Step 2.\(^{25}\) While Step 1 (mother’s latency to hide) did not add a significant amount to \(R^2\), Step 2 (condition) did (\(\Delta R^2 = 0.14, p = 0.04\)). The full regression model (see Table 8) was significant \(F (3, 42) = 3.19, p = 0.03, R^2 = 0.19\), and revealed a significant difference in smile frequency by condition, which was not evident before: Children in the face-to-face condition smiled more than those in the video chat condition \((p < 0.01)\), and those in the video condition also smiled more than those in the video chat condition \((p = 0.01)\). There was no difference between video and face-to-face. The mother’s latency to hide was also a significant predictor \((p = 0.01)\) when entered into the model with condition as another predictor (i.e. Step 2), such that each additional second that it took the mother to begin playing peek-a-boo predicted about 1 additional smile from the baby on average.

Table 8

<table>
<thead>
<tr>
<th>Latency to Hide and Condition as Predictors of Smile Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>------------------</td>
</tr>
<tr>
<td>Constant</td>
</tr>
<tr>
<td>Mom’s Latency to Hide</td>
</tr>
<tr>
<td>Video Chat Condition</td>
</tr>
<tr>
<td>Video Condition</td>
</tr>
<tr>
<td>(R^2)</td>
</tr>
<tr>
<td>(F)</td>
</tr>
</tbody>
</table>

In other words, the test revealed that the mothers’ latency to hide had a positive effect on baby smiles; when this positive influence on baby smiles was accounted for in the model, babies in the video chat condition (where latency to hide was longest) smiled less often than those in the other conditions. Future studies should examine experimentally whether mothers in video chat conditions are successfully buffering the lower number of smiles by engaging with the baby for

\(^{25}\) A third step examining the effects of the interaction terms was also attempted, but this step did not add significantly to \(R^2\).
longer prior to beginning the game. Note that the third step in the regression, which added the interaction terms for condition and latency to hide, neither added significantly to $R^2$ nor created any statistically significant interactions; thus, latency to hide is equally important for producing baby smiles in all three conditions, but mothers in the video chat condition (the condition in which it was most needed) appear to have waited longer to begin playing than those in the other two conditions. This longer latency time in the video chat condition may help explain the null results of Hypothesis 1 (i.e. no differences in smiling across conditions). While we do not know why the mothers’ in the video chat condition took longer to begin playing than those in the other conditions, we can hazard at least one guess: Mothers in the video chat condition may have desired to confirm their babies’ engagement before playing, given that they were interacting using an abnormal, somewhat impoverished (relative to face-to-face) medium; mothers in the video condition may have had similar desires, but with no visual or audio feedback to confirm their babies’ engagement, they would have had no reason to act on this impulse; and mothers in the face-to-face condition may not have sensed a need to spend additional time engaging their child before playing, given that playing face-to-face is a fairly ordinary activity. Alternately, mothers in the video chat condition, who experienced the most abrupt reunion with their children among mothers in the other conditions, may have simply been excited to suddenly encounter their babies after a separation and used more time to re-engage with them. Mothers in the video condition, since they could neither see nor hear their babies, were effectively still in a separation period and thus would not have experienced an exciting reunion. Mothers in the face-to-face condition, while they could not see their babies prior to the start of the trial, were in the same room as their infants for between 30 seconds and 1 minute prior to the trial and thus may have
experienced a more gradual reunion with their children than their peers in the video chat condition, i.e. via hearing their babies or simply by sharing space with them.

The results of the present study suggest several important notions. First, maternal sensitivity is more important for eliciting positive responses from babies than the medium by which they interact. In other words, when it is necessary, mothers can use the same repertoire of engaging interaction techniques that they might use at home to successfully interact with their babies by video chat. This is encouraging, given that some parents rely on video chat interaction with their children when separated from them by work or travel (Yarosh & Abowd, 2011). Second, when mothers spend additional time engaging their babies before launching into more formal play activities, they receive a more positive response from their babies. More research is needed on the relation between this behavior and the babies’ responses; however, if this strategy is shown to be successful in systematically increasing babies’ positive responses to video chat, this behavior could be used as a guideline for parents who are hesitant or unsure of how to interact successfully with their young children via video chat. Instructions like these could be especially useful for parents who may have little experience interacting with their babies in person, like deployed or incarcerated parents. Spending a little bit of extra time at the start of a video call to connect with the baby could play an important role in establishing a positive interaction between babies and their remote family members. Future research should identify other successful behaviors that parents can use to engage their babies during video interactions, with the ultimate goal of educating parents – or even “training” at-risk parents in intervention settings – on how to interact successfully with their children remotely.

Third, there is some evidence here and in Gusella et al. (1998) that babies are sensitive to the difference between their mothers on screen and face-to-face, as evidenced by a longer
hesitation before their first smile. Given this possibility, it is important that one not mistakenly conclude that video chat is “just as good” as face-to-face interaction for babies. While they may not be able to verbalize it, their behavior demonstrates a difference in their perception of these interactions, regardless of their age or experience using video chat in the past. While video chat may provide a promising mode of remote communication for babies and their remote relatives, it’s not the same as traditional interaction. To put it colloquially: Video chat may be better than *nothing* (i.e. no interaction with the absent parent), but it’s not as good as the real thing.

This is especially important to consider under circumstances in which video chat is used to replace face-to-face contact. For example, according to a new report by the Prison Policy Initiative (Rabuy & Wagner, 2015), many jails (but rarely state prisons) have begun implementing programs incorporating video chat visitation between inmates and their families. In a survey of these jails, they found that 74% of them subsequently banned in-person visitation after implementing the video chat visitation programs. In fact, one of the two main companies supplying video chat services to jails and prisons includes this requirement in their contract – i.e. that in-person visitation must be banned following implementation of their for-profit video services. Given the dearth of research on the topic, such actions would appear to be hasty when applied to families with babies. Further research examining both the short- and long-term impact of exclusive video chat interaction between children and parents is urgently needed.

**Limitations and Future Directions**

This study had a number of important limitations. First, there were occasional lags between the mother’s video and audio input in the screen conditions. While there were no differences between the screen conditions in the number of sessions that included this error, we were not able to measure the degree of the delay (i.e. whether some delays were longer than
others) in this study. Delays like these occur frequently during naturally occurring video chat interactions (Chapter III, this dissertation), so it can be argued that there is some ecological validity to this unintended error; however, it does dampen the degree to which these tests can measure differences due to condition alone. Future studies should incorporate methods of measuring unintended lags so that this variable may be used as a covariate in analyses.

Similarly, mothers in the screen conditions occasionally disappeared from view unintentionally during the study. Relatives are frequently and unintentionally obscured from view when using video chat at home (McClure & Barr, *under review*), which again suggests that this unintended error has some ecological validity; however, again, it does produce a limitation in this otherwise well-controlled experimental study. Fortunately there were no statistically significant systematic differences between the two screen conditions on this dimension, and using it as a covariate made no difference to the outcome of the tests. Future studies should provide clear instructions to mothers regarding the extent of their visibility via the webcam.

Second, due to an equipment malfunction during data collection, the skin conductance data produced by babies varied unusually between subjects. The technical problems with this data make it unclear whether the skin conductance data is reliable. Future studies should include this measure and investigate whether the skin conductance results in this study can be replicated. Furthermore, respiratory sinus arrhythmia (RSA) is a powerful tool used in studies of attention to and engagement in digital media presentations, and its use has been encouraged by media researchers (Ravaja, 2004). Future studies of this kind should include this measure.

Third, unlike Hains & Muir’s study using strangers (1996), we used a between-subjects design. In one sense, this is a strength of the study, as it adds a novel approach to the existing literature, which has relied entirely on within-subjects designs. Furthermore, given the addition
of a face-to-face condition in this study, a within-subjects design would have required a rapid mother-infant separation following the face-to-face interaction in order to move to the subsequent condition, which would almost certainly create infant anxiety and fussiness. Given the serious consequences of producing such anxiety during a study measuring infant emotional responsiveness, the current design was the stronger of the two options; however, as with all between-subjects designs, the results should be interpreted carefully.

**Conclusion**

This experimental study was the first to systematically compare infants’ emotional responsiveness to their mothers across face-to-face, video chat, and non-contingent video. Prior to this study, it was unclear how sensitive, face-to-face mother-infant interactions would be affected by screen mediation.

The results reveal that the game of peek-a-boo is equally exciting for babies when their mothers play it with them, regardless of the medium chosen. Furthermore, the babies’ positive responses were more dependent on their mothers’ sensitivity than on the mode of interaction. This suggests good news for mothers who rely on video chat to interact with their babies when they are separated from them: it appears that mothers can use their usual repertoire of sensitive interactions and games to elicit positive responses from their babies, even via video chat and pre-recorded video. It also provides support for the use of programs like *United through Reading*, (Yeary, Zoll, & Reschke, 2012), a program that allows deployed parents to create video recordings of themselves reading a children’s book aloud to their child at home. This study only demonstrates short-term tolerance of non-contingency in screen presentations, though, so the use of video chat for geographically separated parents and infant may also be important.
Finally, the results support previous work demonstrating that babies are sensitive to differences between video chat and face-to-face interactions on at least one dimension. (i.e. latency to smile) – even in the unnatural face-to-face scenario produced in the lab. For this reason, and because the research in this area is still quite limited, face-to-face parent-child interactions should still be considered the standard by which all mediated interactions are judged.
Chapter V: General Discussion

The present dissertation, in the form of three studies, explored the role that video chat technology plays in facilitating communication between babies and their distant family members. In the first study (Chapter II), an electronic media usage survey was utilized to assess the degree to which D.C.-area families with children under 2 currently use video chat. This study demonstrated that the use of video chat among infants and toddlers is remarkably common, even among those whose parents know and otherwise follow the AAP’s media usage guidelines for children in this age range. In fact, parents who placed restrictions on, or whose child otherwise had minimal exposure to, other forms of media exempted video chat from their limited screen use. Many families in the sample reported using video chat with their babies one or more times per week; given that observed video calls have an average length of about 20 minutes (Chapter III), this is a significant and previously unreported source of media exposure for children in this age range.

In the second study (Chapter III), a naturalistic observational method was employed to examine the way families use video chat at home with their children under 2 years of age. Overall, the interactions were surprisingly successful, involving high levels of shared attention, pretend play, and physical affection, despite technological and developmental barriers to their use and the frequent occurrence of technical problems.

In the third study (Chapter IV), a controlled experimental design was used to systematically compare the emotional engagement of 6- to 12-month-olds during mother-baby interactions taking place either face-to-face, via video chat, or via non-contingent video. Three important findings surfaced. First, babies appeared to be fairly tolerant of screen presentations of their mothers, whether contingent or not, as demonstrated by their equal levels of emotional
reactivity across the three conditions. Furthermore, the mothers in the video chat condition spent longer engaging their babies prior to beginning the peek-a-boo game, a behavior that counterbalanced their babies’ tendency to otherwise smile less often than those in the other conditions. Second, there was some evidence that babies were emotionally sensitive to the difference between their mothers face-to-face and by video chat, as demonstrated in a longer delay before their first smile in the video chat condition. Third, maternal sensitivity during the interaction played a more significant role in eliciting positive responses from the babies than did the mode of communication (i.e. experimental condition).

This dissertation adds to the current literature in a number of ways. It is the first to document the degree to which families with babies are currently using video chat at home, thus adding to our understanding of babies’ media landscape (Rideout, 2013). It is also the first to observe families with children under two using this technology in their homes, adding to previous observational work with older children (Ballagas, Kaya, Ames, Go, & Raffle, 2009). It is also the first to compare mother-infant interactions across all three conditions – face-to-face, video chat, and non-contingent video – in a single study, and to add psychophysiology as an additional measure of emotional reactivity in babies. This contributes valuable information and methodological strength to the current literature (e.g. Hains & Muir, 1996).

**Theoretical Implications**

The results of these studies have important theoretical implications. Babies today are immersed in a digital world, in which formative relationships with important family members like grandparents are often mediated by a screen. It is possible that well-established developmental processes, including the development of ToM, JVA, JPP, and social-emotional communication skills are now being influenced by frequent, lengthy exposure to video chat.
interactions. The results of these studies indicate that researchers should consider media exposure when studying these processes, and suggest a promising new avenue of developmental research.

First, the development of JPP may be influenced by exposure to video chat. For example, in the home observations described in Chapter III, toddlers were observed giving “screen kisses” to their grandparents. While some behaviors like these – in which the child behaves as though the events on screen are real (i.e. are capable of being interacted with physically) – have been observed in prior research with non-contingent videos (Pierrotsakos & Troseth, 2003), the behaviors seen here took place in an important new context: a context in which such behaviors are contingently responded to by the person on screen and are encouraged as though they are real. This leads to an important new question: Are such behaviors “real” or “pretend” in the context of video chat? Do the toddlers’ attempts to physically interact with their remote grandparents signify an underlying misunderstanding of the medium, or do these behaviors support true relationship development between the child and relative? Unlike prior research, in which babies’ and toddlers’ attempts to respond to screen stimuli as though they were real were called “fruitless” (Pierrotsakos & Troseth, 2003, p.196), the behaviors seen here may be considered by the video chat participants (including the adults) to be successful and actual transfers of affection or contact. If the adults in this context act as though these interactions are real (and may even believe that they are), how will this influence the child’s developing understanding of what is real and what is pretend, what is a depicted screen image of Grandma and what is real Grandma? Are parents establishing a new culture of what is real as they encourage their babies to “kiss” Grandma or give Grandpa a high five, and as they consistently behave as though the screen depiction of these relatives is the real person? Will young children
outgrow these behaviors as their knowledge of the relationship between depicted referents and their representations develops (Pierroutsakos & Troseth, 2003), or will these behaviors persist, falling into some other ambiguous category— not as simple misunderstandings, but as real-yet-remote affectionate interactions? While the child who said her grandpa lived “right there” in the iPad probably misunderstands her grandfather’s true living situation, she also has a point: her relationship with her grandfather really does live and take place right there (probably more often than not), through the screen before her.

Second, as discussed more extensively in Chapter III, the development of ToM may affect and be affected by children’s video chat experiences. Future research should pursue the relationship between ToM development and: the ability of a child to intentionally and successfully manipulate the camera’s viewpoint during video chat; a child’s video chat exposure; and the role of video chat as a delivery method (versus non-contingent video) of demonstrations for ToM assessments.

Third, video chat usage and JVA development mutually influence one another in important ways. The use of across-screen JVA predicted babies’ attention during the video chat calls in their homes, which is not surprising given that the purpose of across-screen JVA is to direct attention to and across the screen. More interesting is the fact that a clear developmental trend was demonstrated among toddlers across the types of JVA available to them during video chat: Only younger babies did not initiate any JVA at all; both younger and older babies initiated traditional (“within-screen”) JVA to reference objects in their own environments; but across-screen JVA, a complex and novel form of JVA for children, was initiated almost exclusively by the older babies in the home sample. Across-screen JVA, a term coined in this dissertation, is a new and previously-uninvestigated phenomenon among very young children. The home visit
study established its temporal location within the developmental timeframe of JVA as a whole, where across-screen JVA was found to develop last among the types of JVA observed. Future research should establish whether this is due to its complexity or to the fewer opportunities babies have to practice it (or both).

Fourth, the results of these studies have important implications for the video deficit (a.k.a. transfer deficit) effect. Very young children, roughly between 6 months and 3 years of age, exhibit a remarkable deficit in learning from video compared to live demonstrations – a phenomenon that has come to be known as “the video deficit effect” (Anderson & Pempek, 2005). For example, young children who are asked to imitate a task demonstrated on screen do significantly worse on the task, on average, than children asked to imitate the same task demonstrated by a live model (Barr & Hayne, 1999). The deficit appears across many task and content types (Krcmar, Grela, & Lin, 2007; Kuhl, Tsao, & Liu, 2003; Troseth & DeLoache, 1998), suggesting a fairly broad deficit in early learning from screens. There is evidence, however, that this transfer deficit can be entirely overcome when children are exposed to a live video feed prior to testing, possibly because it helps them establish that the events on screen are “real,” and thus relevant and applicable to their lives. When 2-year-olds had access to a live video feed for 2 weeks in their own homes, 2-year-olds were able to overcome the video deficit in a lab-based object-retrieval task (Troseth, 2003). Furthermore, Troseth, Saylor, and Archer (2006) found that just 5 minutes of exposure to a live video feed immediately prior to testing was enough to extinguish the deficit among 2-year-olds. These studies support Troseth’s (2010) proposal that children of this age do not typically view video screens as a source of information relevant to their lives unless they are persuaded otherwise. She suggested that infants, who are exposed early and often to television (Rideout, 2013), quickly discover that screens do not
necessarily portray present reality and that the people portrayed on them are not true social partners. By distinguishing events on screens as “not real,” infants would then be likely to struggle when asked to apply information from screens, now distinguished as symbolic, to their “real” lives, a process that requires cognitively-taxing dual representational thought – or the ability to simultaneously hold in mind both a symbol and its referent (DeLoache, 2000).

According to this thinking, young children should be more likely to learn screen content if experimenters can extinguish toddlers’ belief in the separateness of screen presentations from reality, an hypothesis supported by Troseth’s (2003; Troseth et al., 2006) studies.

Will the video deficit continue to exist, then, given the high rates of video chat exposure among babies and toddlers (Chapter II)? While the current evidence certainly suggests that the video deficit should be mitigated by this exposure, there is at least one potential alternative. It is possible that toddlers with high levels of video chat experience will become experts at distinguishing between video chat and other screen stimuli, and will thus recognize when screen stimuli are (or are not) “socially relevant”. In other words, it’s possible that a little video chat exposure – like the exposure introduced in Troseth’s (2003; Troseth et al., 2006) studies – will lead children to see all screens as potential sources of socially relevant information, but that extensive video chat exposure will lead children to ignore non-contingent video information because they can quickly recognize that the source is not video chat. If this is the case, there will be a U-shaped curve in the relationship between video chat exposure and the video deficit. Replications of the original video deficit studies, which should take into account each child’s previous exposure to video chat, are necessary to determine whether this phenomenon still exists and, if so, what its relationship to both video chat experience and age will be.
Fifth, the home visit observations add to our understanding of coviewing and scaffolding behavior (Wood, Bruner, & Ross, 1976) in the context of modern media. Previous studies on parent coviewing of TV have shown that children learn more from screens when a parent coviews with them (for a review see Takeuchi & Stevens, 2011). While a shared adult-child media experience occurs in the context of video chat as well, the experience is not adequately described by the word “coviewing”: instead of simply sharing attention to screen content that is delivered relatively passively to them as viewers, here both the parent and the child are collaboratively involved in engaging with the screen and interacting live with the person on it, navigating challenges together. Rather than calling this coviewing, we might instead call it very early joint media engagement (JME), a newer and more flexible term that refers to any kind of media co-use by multiple individuals (Takeuchi & Stevens, 2011) and that can encompass the adults’ powerful scaffolding techniques as they co-use technology with their children. In fact, the adults in the observed video chat calls were not simply co-users with the children; they were co-learners. In other words, parents were not simply experts scaffolding the novice child’s understanding (Vygotsky, 1978) of screen content; instead, while scaffolding the child’s complex interaction with the person on the screen, they themselves often encountered obstacles – sometimes the very same obstacles encountered by the child – that challenged their own understanding of the technology. For example, across-screen JVA is challenging for parents, grandparents, and children alike: it’s not always clear even to adults what the remote party is able to see, a challenge that parents and grandparents negotiated by verbally advising one another. In such instances, parents were involved in scaffolding for their young child a process that they themselves were still learning and that required the assistance of others to resolve. In fact, unlike some skills in which the scaffolds may be removed once a learner has mastered the
skillset and can accomplish the task unaided, complex mediated phenomena like across-screen JVA may require the occasional support of other co-participants even after the skill is “mastered”. This complex, intergenerational support system for collaborative JME is worthy of additional study.

Finally, these studies have implications for our understanding of young children’s responses to sudden interruptions in responsiveness during social interactions. The extensive existing literature on the still-face effect demonstrates that infants as young as 3 months of age show decreases in smiling and avert their gaze when their mothers abruptly stop responding to them, holding a still face, after a period of continuous interaction (for a review and meta-analysis of the effect see Mesman, van Ijzendoorn, & Bakermans-Kranenburg, 2009). Why, then, did technological delays – which often suddenly produce a frozen image of the screen partner – not have a significant negative effect on child outcomes in these studies? There are a few potential explanations. The first, as discussed in Chapter III, is that our measure of delay was too inexact to capture the frequency of true delays in the natural home video chat interactions. In this case, we should be very cautious about interpreting the result, as it may underestimate the importance of naturally occurring delays for children. The second, also discussed in Chapter III, is that children did not respond negatively to the delays at home because their parents model and scaffold positive responses to these problems, like reframing a technological problem as a suspenseful game. This fits with previous research demonstrating that infants do not react negatively to sudden adult non-responsiveness when there is a clear reason for it, as when the adult looks away to speak with someone else (Mesman, van Ijzendoorn, & Bakermans-Kranenburg, 2009). By reframing technological problems as purposeful and modeling positive responses, parents may be able to decrease (or remove entirely) their disturbing quality for
babies; however more research is needed to determine whether this is the case. The still-face paradigm has been studied almost exclusively in laboratory settings, where the baby is left alone under highly controlled conditions. The technical problems that occur during home video chat usage provide a naturally-occurring complement to the still-face paradigm, which creates important opportunities for further investigations: Future research should explore this naturally-occurring still-face scenario in more depth within the home setting, investigating the differential effect on babies and toddlers when parents naturally either do or do not reframe technological problems for their babies when they co-use video chat.

**Practical Implications**

The results of these studies have important practical implications. First, while the research is still very limited, these findings suggest that video chat usage is an exceptional type of media exposure. The AAP should take note of this as they update their media usage guidelines for children under two. At the very least, parents appear to perceive it as a more positive type of screen exposure than other kinds, worthy of exception to their ordinary media restrictions (Chapter II). Unlike television and videos, video chat is a socially contingent form of media use, which successfully engages the child in the kinds of sensitive social interactions that are important for their early development (Stayton & Ainsworth, 1973; Chapter IV). Furthermore, unlike television and videos, families with babies do not engage in video chat usage for the sake of entertainment or education; instead, they use it to maintain social bonds with remote relatives like grandparents (Chapters II & III), an endeavor that, if successful, may have the ultimate effect of increasing social support among the family unit and providing access to the positive influence of grandparents in the lives of their grandchildren (Dunifon, 2013). The AAP’s current recommendations are based primarily on research examining television exposure, and they
provide a conservative suggestion based on the scarcity of research showing positive effects (e.g. learning) from exposure and an established literature showing the potential for negative effects (e.g. on attention) from long-term, high levels of it (AAP, 2011). Television research may not be applicable to other types of modern media use, and a more nuanced set of guidelines would be beneficial for families. The AAP should consider the potential for positive outcomes from other media forms – video chat in particular – when the recommendations are revised in the coming year (Brown, Shifrin, & Hill, 2015).

This work as a whole also has important implications for parent education and intervention programs. Findings across the studies could be used to educate and instruct families in the best practices for interacting with babies and toddlers via video chat. For example, findings from the survey study (Chapter II) suggest that it may be beneficial for families to try using an open video connection on certain occasions, like holidays and birthdays. The surveyed families reported that it increased the feeling that the remote relative was “really there” with them in the room, a concept termed “social presence” in the human-computer interaction literature (Biocca, Harms, & Brigoon, 2003). Instead of using video chat as a way to have a focused interaction with the remote relative exclusively, using an open video connection allows families to simply include the relative in the general milieu of an event – whether that be a family meal, a birthday party, or a holiday celebration – as a present member of the community. It also has the potential to allow participants to spend time with one another, while reducing the pressure to maintain a verbal conversation or take part in long, focused interactions that may be difficult for young children.

26 Many of these studies are limited and have been criticized by others in the field. For more on this, see Lerner & Barr, 2014.
Furthermore, in the home study (Chapter III), it became clear that the present parents played a very powerful role as facilitators of the interaction. Parents can be instructed to use successful scaffolding techniques (e.g. helping children interpret and initiate across-screen JVA; using mind-minded language to facilitate ToM-supported activities using the camera; modeling positive, game-like responses to technical problems and delays) and to participate in grandparent-child activities as both facilitators (e.g. helping to maintain the child’s attention) and as physical proxies for the grandparent (e.g. acting in the place of a grandparent during the “This Little Piggy” game; squeezing the child when the grandparent gives a screen hug). Finally, results from the lab study (Chapter IV) suggest that children will be more emotionally responsive in video interactions if parents and grandparents spend a few minutes at the beginning of the call to engage the child and secure their attention prior to initiating formal play activities.

Suggestions like these may be useful to all grandparents and parents who use video chat, but they will be especially beneficial for remote relatives who are under stress (e.g. who are incarcerated, deployed, or facing work demands), are inexperienced interacting with children (e.g. teens, some fathers), or are remote from the children for longer periods (e.g. remote grandparents and incarcerated or deployed parents). Guidelines and training modules for these populations have the potential to be very helpful.

These studies also have important implications for developers responsible for creating video chat software and apps. Given the potential of this medium for young children and the prevalence of its use by them, designers need to keep the needs of babies, toddlers, and very young children in mind as they consider the features of their software. For example, very young children tend to make frequent attempts to touch the screen while using video chat, which are discouraged by their parents. This may have adverse consequences for the call’s success,
including accidental call termination, frustration among the adults, frustration for the child whose attempts are discouraged, and possibly to foreshortened calls. Given the importance of physical contact for children’s social interactions, designers should consider creative ways to allow children to touch the screen during video calls to allow them to express affection, play physical games with their relatives, and interact more freely with the device, all while avoiding call termination. While some touchscreen devices have device settings that allow users to turn off touch sensitivity (e.g. iPhone’s Guided Access allows users to “turn off” touchscreen sensitivity for user-selected parts of the screen), many parents are unaware of these tools and they can be unwieldy to deploy successfully, especially for parents who are busy with their toddlers while setting up a call. Developers could integrate this technology into existing video chat apps as a “child-proof setting,” which would both alert parents to the existence of this tool and make its application within the session easy. Combined with the use of a heavy-duty device case, these tools could make video chatting a more child-friendly experience.

App developers also have the opportunity to incorporate scaffolding for children directly into their products. For example, Kindoma has created several video chat apps for preschoolers that allow users to participate in shared activities on the screen: the Storytime app (Kindoma, 2013) allows remote relatives and children to co-read a children’s e-book, while still seeing one another’s faces and also supporting JVA by visually indicating on-screen pointing between the partners (see Figure 11); and the Drawtime app (Kindoma, 2015) allows children and their remote relatives to draw pictures together on the screen.
These products support JVA and shared activities between children and remote adults, both of which played important roles in the success of calls in the home study. These apps were created with the needs of 3-year-olds in mind; however, future apps should also consider the special needs and interests of babies and toddlers, perhaps simulating or facilitating activities like the transfer of screen affection or the sharing of food and play objects across the screen.

Conclusion

The findings as a whole suggest that video chat is a promising mode of remote communication for families with babies and toddlers: it is accessible to families and is actually used by them; when it is used, it is done fairly successfully; and the kinds of sensitive behaviors that are used with babies face-to-face can also be used to engage positively with babies via video chat. Telephone use has proven to be an unsuccessful mode of communication for babies and toddlers (Tarasuk, Galligan, & Kaufman, 2013; Ballagas et al., 2009), a finding that is consistent with its low usage among families in Chapter I. Some researchers have suggested video chat as a

---

possible alternative for remote communication among very young children (Ballagas et al., 2009), and these studies support that proposal.

The three studies in this dissertation provide a novel examination of infant and toddler engagement in video mediated interactions. Given their novelty, future research attempting to replicate and extend these findings is called for. Many families, both immediate and extended, are geographically separated from one another today. By studying the role that video chat may play in bridging the geographical divide between babies and their remote family members, researchers can provide information to guide families, developers, and policy makers as they determine the potential value of this new technology.
Appendix A
Code Definitions and Reliability

Displays of Physical Affection
Active, non-habitual displays of physical affection between partners (e.g. kisses, hugs).
Utilitarian touches are excluded (e.g. fixing the child’s hair), as is the resting position of the baby (e.g. sitting on an adult’s lap may be considered a display of physical affection in some cases, but not if the adult is simply acting as a passive place for the child to sit).

- Reliability: ICC = 0.94

Touching the Device
Touching or attempting to touch the screen or device. This is the discrete instant in which a touch (or touch attempt) is made, not the duration during which it is touched. Includes moments when a device changes hands.

- Successful Attempt
  An attempt to touch the device is successful; the individual makes contact with the device.
  - Reliability: ICC = 0.97

- Unsuccessful Attempt
  An attempt to touch the device is unsuccessful; the individual does not make contact with the device.
  - There were not enough instances to test for reliability. This code was dropped from all further analyses.

- Encouragement of the Attempt
  An individual other than the one making the attempt verbally encourages the attempt (e.g. “Give Grandma a hug!” and “Can you press the ‘end call’ button?”).
  - Reliability: ICC = 0.79

- Discouragement of the Attempt
  An individual other than the one making the attempt verbally discourages the attempt (e.g. “No, don’t touch! You’ll hang up on her!”).
  - Reliability: ICC = 0.93

- Neutral Attitude toward the Attempt
  Neither encouragements nor discouragements of the touch attempt are verbally made.
Joint Visual Attention (JVA)

One individual directs another individual’s attention to a third object, person, or event.

- **Within-screen JVA**
  The JVA initiator directs a partner’s attention to an object or person on the same side of the screen (e.g. a baby brings a toy to the screen to show the grandparent).
  - Reliability: ICC = 0.77

- **Across-screen JVA**
  The JVA initiator directs a partner’s attention to an object or person on the other side of the screen (e.g. the baby points to her grandmother’s dog).
  - Reliability: ICC = 0.91

- **Within-beyond screen JVA**
  The JVA initiator directs a partner’s attention to an object or person on the same side of the screen that is beyond the view of their partner (e.g. the baby points out the window to a passing truck, with the grandparent cannot see).
  - There were not enough instances to test for reliability. This code was dropped from all further analyses.

- **Across-beyond screen JVA**
  The JVA initiator directs a partner’s attention to an object or person on the other side of the screen that is beyond their own view (e.g. the baby asks to see the grandparent’s dog, who is currently off-screen).
  - Reliability: ICC = 1.00

- **Successful JVA**
  The person whose attention is being directed directs their attention to the object or person intended by the initiator.
  - Reliability: ICC = 0.70

- **Unsuccessful JVA**
  The person whose attention is being directed does not direct their attention to the object or person intended by the initiator.
  - Reliability: ICC = 0.82
Joint Pretend Play (JPP)

The participants engage in a social activity that involves imaginary or impossible objects, actions, or ideas. This includes trying to pass objects through the screen.

- **Within-screen JPP**
  
  Joint Pretend Play between two or more individuals that takes place on only one side of the screen. The Play initiator is on the same side of the screen as the individual invited to play (e.g. the baby pretends to share a cup of tea with her mother).

  - Reliability: ICC = 1.00

- **Across-screen JPP**
  
  Joint Pretend Play between two or more individuals that takes place across the screen. The Play initiator is not on the same side of the screen as the individual invited to play (e.g. the baby pretends to share a cup of tea with her remote grandmother).

  - Reliability: ICC = 1.00.

- **Successful JPP**
  
  The individual invited to play reciprocates by entering into the shared play activity (e.g. the grandmother accepts the pretend cup of tea from her grandchild and pretends to drink it).

  - Reliability: ICC = 1.00.

- **Unsuccessful JPP**
  
  The individual invited to play does not enter into the shared play activity (e.g. the grandmother does not notice the baby offering the pretend tea cup and continues talking about other things).

  - Reliability: ICC = 1.00.

**Attention**

- The amount of time the child spends in screen-directed behaviors (e.g. gazing, gesticulating, vocalizing, showing things, or participating in activities toward the device). Reliability: ICC = 0.91
Technological Talk

- The amount of time participants spend talking about the technology they are using for the video call, including talking about any technical difficulties. Reliability: ICC = 0.99

- Technological Talk about Seeing
  The technological talk regards any individual’s ability or inability to see a partner or partners through the camera’s view (e.g. “Can he see me?” and “We only see part of your face!”). This includes requests to adjust the camera’s position or angle.
  - Reliability: ICC = 1.00

Technological Problems

- The amount of time spent in technological problems that involve significant visual, audio, or visual-audio delay; also problems leading to call termination. Durations end when the problem is completely resolved, either within the same call period or at the restart of a dropped call. Reliability: ICC = 0.96

- Problem Source: The technology itself
  The technology itself is responsible for the technological problem. This includes delays due to slow connections.
  - Reliability: ICC = 0.99

- Problem Source: The baby
  The baby causes the technological problem. This includes instances where the baby accidentally or intentionally hangs up the call by touching the device.
  - Reliability: ICC = 0.99

- Problem Source: The adults
  An adult causes the technological problem. This includes instances in which the audio or visual input is not present because an adult does not know how to start them.
  - There were not enough instances to test for reliability. This code was dropped from all further analyses.
Center of Interaction
A scale of 1 to 3, rating whether the call was: mostly child-centered (1), equally child-centered and adult-centered (2), or mostly adult-centered (3).

• There was only one rating for each call in its entirety; thus, reliability was assessed across calls, rather than within calls.
  ○ Reliability: ICC = 1.00

Interaction- or Observation-Based Call
A scale of 1 to 3, rating whether the call was based: mostly in interaction with the child (1), equally in interaction with and observation of the child (2), or mostly in observation of the child (3).

• There was only one rating for each call in its entirety; thus, reliability was assessed across calls, rather than within calls.
  ○ Reliability: ICC = 0.95

Baby Activity Level
A scale of 1 to 4, rating the activity level of the child; specifically, how often the child moves from place to place: Stationary (1) – The child does not change locations at all; Low (2) – The child changes locations very rarely; Moderate (3) – The child changes locations regularly; High (4) – The child changes locations nearly constantly.

• There was only one rating for each call in its entirety; thus, reliability was assessed across calls, rather than within calls.
  ○ Reliability: ICC = 0.91

Device Movement
A scale of 1 to 4, rating the frequency with which the device is moved or transferred throughout the interaction: Stationary (1) – The device does not move at all; Low (2) – The device moves very rarely; Moderate (3) – The device moves regularly; High (4) – The device moves nearly constantly.

• There was only one rating for each call in its entirety; thus, reliability was assessed across calls, rather than within calls.
Adult Sensitivity

The sensitivity of the adult participants in the call was assessed using the Parental Sensitivity Scale from The Emotional Availability Scales, abridged Infancy/Early Childhood Version (Birignen, Robinson, & Emde, 2000). Continuous ratings on this 1-9 scale were recorded using a joystick, and the average value for each adult in each call was used for analysis and reliability assessment.

- **Primary Present Adult**
  
  Coders selected the primary present adult for sensitivity assessment. If two adults were present, the primary present adult was defined as the present adult who was more involved in the video call.
  
  - Reliability: ICC = 0.91

- **Primary Remote Adult**
  
  Coders selected the primary remote adult for sensitivity assessment. If there were two adults on the screen, the primary remote adult was defined as the remote adult who was more involved in the video call.
  
  - Reliability: ICC = 0.92
## Appendix B
### Attention Correlation Table

**Correlations between six variables and attention**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Attention</td>
<td>0.27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Parent's sensitivity</td>
<td></td>
<td>0.20</td>
<td>0.12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Frequency of baby affection</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Frequency of baby device-touch attempts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Discouragement of device touching</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Frequency of technical problems</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>JPP with baby</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p ≤ 0.05; **p ≤ 0.01; ***p ≤ 0.001
References


doi:10.1016/j.earlhumdev.2007.09.019


During the Second Year. *Journal of cognition and development, 8*, 401-425. doi: 10.1080/15248370701612951


