Enhancing Technology-Mediated Communication: Tools, Analyses, and Predictive Models

Daniel Avrahami
דניאל אברהמי

September 2007
CMU-HCII-07-102

Human-Computer Interaction Institute
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Thesis Committee:
Scott E. Hudson (Chair), Carnegie Mellon University
Susan Fussell, Carnegie Mellon University
Robert Kraut, Carnegie Mellon University
Eric Horvitz, Microsoft Research

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

This work was supported by a Graduate Fellowship from Carnegie Mellon University, and supported in part by the National Science Foundation under grants IIS-9800597, IIS 0121560 and IIS 0325351, and by the Defense Advanced Research Projects Agency (DARPA) under Contract No. NBCHD030010. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author and do not necessarily reflect those of the funding agencies.

Keywords: Interpersonal communication, technology-mediated communication, computer-supported cooperative work, CSCW, computer-mediated communication, CMC, interruption, responsiveness, availability, interpersonal relationships, predictive statistical models, task-switching.

Abstract

For most of us, interpersonal communication is at the center of our professional and personal lives. With the growing distribution of business organizations and of our social networks, so grows the need for and use of communication technologies. Many of today's communication tools, however, suffer from a number of shortcomings. For example, the inherent discrepancy between one's desire to initiate communication and another's ability or desire to receive it, often leads to unwanted interruptions on the one hand, or failed communication on the other. I have taken an interdisciplinary approach to address these shortcomings, and also in order to provide a better understanding of human behavior and the use of communication tools, combining tool-building and the creation of predictive models, with investigation and analysis of large volumes of field data.

At the focus of this dissertation is my research on Instant Messaging (IM) communication, a popular, interesting, and highly observable point on the continuum between synchronous and asynchronous communication mediums. I present the creation of a set of statistical models that are able to predict, with high accuracy, users' responsiveness to incoming communication. A quantitative analysis complements these models by revealing major factors that influence responsiveness, illuminating its role in IM communication. I then describe an investigation of the effect of interpersonal relationships on communication, and statistical models that can predict these relationships. Finally, I describe a tool I have created that allows users to balance their responsiveness to IM with their ability to stay on task.

Acknowledgements

I will start by thanking my advisor Scott, for mentoring me throughout my time in graduate school. Scott has taught me many things about research and being a researcher. He taught me about academia, and taught me about teaching. Finally, I thank him for showing patience when I was searching for my own research path. His advising style, which I was lucky to experience, comes directly from his being a good and kind person. I have enjoyed working with him over the past years and I hope that we will be able to collaborate again in the future. I also thank my other committee members, Robert Kraut, Eric Horvitz, and Sue Fussell for comments and suggestions that have undoubtedly improved this dissertation. I am indebted to Sue Fussell. When Scott was on sabbatical and I needed help, she was always willing to spend her time, providing me with support and great advice. Thank you Sue.

To my mother Judith Avrahami and my step-father Yaakov Kareev, I owe my interest in research and my pursuit of graduate studies (yes, I also owe them for their help with statistics). By introducing me to computers and to cognitive psychology they can be credited (or blamed) for my belonging to the field of Human-Computer Interaction.

A big thank you to Darren Gergle. First and foremost for his friendship -- he played a big part in making graduate school a fun and memorable experience. I should also thank Darren for his advice and help with the work and writing, and for his words of encouragement when I (as one is bound to) got tired of the process. We should write another journal paper together so it can be printed when his first grandchild is born.

To the HCII faculty and staff for providing an engaging and challenging research environment which I was glad to be part of. Acknowledging that experts may be located at other departments across campus, encouraging interdisciplinary collaborations, is important and admirable. A special thank you goes to Marian D'Amico. Had it not been for her competence and swift communication, I would have gone to a different university, and for that I thank her.

Many others have helped me with this work, either directly, through good advice, or through suggestions on unfinished manuscripts. These include my friends, peers, and future colleagues Aaron Bauer, Daniel Boyarski, Laura Dabbish, James Fogarty, Jodi Forlizzi, Gary Hsieh, Sara Kiesler, Andy Ko, Jennifer Lai, Alon Lavie, Johnny Lee, Joonhwan Lee, Jeff Nichols, Roni Rosenfeld, Mike Terry, Irina Shklovski, and Jake Wobbrock. Thank you to Anind Day and Jen Mankoff. Their comments on the presentation of this work were instrumental in my pursuit of a research position.

This long process wouldn't have been possible without the support of many good friends. In particular, thank you Carl DiSalvo, Gillian Hayes, Julie Kientz, Yarden Kedar and the clan, Guy Itzhaki, Yoav Etsion, Jess Bauer, Peter Scupelli, Kelly Scupelli, Joe Tullio, Ian Li, and Paul Burkhead.

A big thank you to my family. Ari, Frieda, Ruth, Emanuel, Doda Rachel, Safta Elise, Safta Riva, Saba Adi, Daniella, Moshe, Soroko, Noa, and Ami. I love you all very much.

Finally, a thank you to my wife Thi. For her support, for her patience, for her comments and suggestions (always right), for making me happy, for her love, and for Noam.

To my family

באהבה

Table of Contents

List of Figures

List of Tables

Chapter One

Introduction

Consider the following scenario; Anne is making final changes to a presentation for a client's visit. Her team member John, working at a different site, tries to contact Anne to discuss an urgent issue. However, since Anne is pressed for time, and having already been disrupted a number of times, she has decided to ignore all incoming communication until after she's done, leaving John unable to finish his task.

Consider now an intelligent system that is able to accurately predict, based on her activity, that Anne is not likely to respond to John for some time. A system that is also able to predict, based on past communication patterns, that Anne and John are co-workers, and is able to estimate the urgency of John's request. Such a system would be able, for example, to increase the salience of an alert, indicating to Anne that, among her incoming communication, John's request may deserve her immediate attention. Alternatively, a system could direct John's query to another co-worker who could provide him with a timely response. This document describes the development of tools and models necessary for the creation of such intelligent systems.

The two main goals of my thesis work are:

- Provide a better understanding of factors affecting technology-mediated communication in its context, and
- Use this understanding for the creation of predictive statistical models and tools that can enhance communication.

Focusing my dissertation work on Instant Messaging (IM) communication – a popular, interesting, and highly observable point on the continuum between synchronous and asynchronous communication mediums – I have taken three complementary steps looking at key aspects of communication:

- Investigated the factors that affect responsiveness to IM communication and created models that accurately predict responsiveness to incoming IM.
- Investigated the effect of interpersonal relationships on IM interaction, and created statistical models that use this knowledge to predict relationships.
- Made use of basic properties of human dialogue to create a tool that provides support for balancing responsiveness and performance.

This work's contribution to the HCI field spans both theoretical and applied aspects. From a theoretical point of view, this work advances previous work by providing insights into the factors that influence interpersonal communication patterns and responsiveness. At the applied level, this work provides predictive statistical models that can be used in many useful applications. Finally, this work promotes the creation of tools that use predictive models that are generated from naturally occurring interaction.

1.1 Dissertation Outline

The remainder of this dissertation is organized as follows:

Chapter 2 presents the background to my dissertation work with a review of related literature. I describe, for example, the importance of communication and the link between technology-mediated communication and interruptions.

Chapter 3 describes the process of creation of statistical models that are able to predict, with high accuracy, a user's responsiveness to incoming instant messages. This chapter includes the description of the data-collection mechanism that I created and the recorded data used for the work presented in Chapters 3 through 6.

Chapter 4 describes an examination of the interaction between the time that has passed since the arrival of a message and the likelihood of a response. Unlike the models presented in Chapter 3, which aim to provide benefit through predictions of responsiveness *prior* to the delivery of a message, this chapter examines responsiveness *after* a message has been sent and while the sender is waiting for a response.

Chapter 5 presents an in-depth quantitative analysis of responsiveness. In this chapter I describe the effects of a user's *context,* elements of the *communication*, and features of *content* on responsiveness. Through this analysis I am able to advance our understanding of responsiveness and its relationship with a user's availability.

Chapter 6 describes an investigation of the effects of the relationship between IM communication partners on basic features of their communication. This work extends prior research on the effects of relationship on face-to-face and phone communication. This chapter then presents the use of the findings for the creation of statistical models that classify the relationship between IM users.

Chapter 7 presents a tool that allows users to balance their performance on ongoing tasks with their responsiveness to incoming messages. Specifically, this tool helps users distinguish between messages that require fast responses and those that they are waiting for from others. This chapter is concluded with a preliminary evaluation suggesting the effect of this tool on responsiveness.

Chapter 8 concludes this dissertation by highlighting some of the major findings presented in this document and by pointing to several interesting areas for future work.

Chapter Two

Background

Interpersonal communication is a major component of our personal and professional lives. Indeed, communication is a central activity in most organizations with unplanned, spontaneous communication important for collaboration and the successful completion of work. As the distribution of business organizations and of our social networks increases, the need for and use of communication technologies grows. However, it was previously shown that as physical distance grows, communication and collaboration decreases (Kraut, Fish, Root, & Chalfonte, 1990). In particular, when communication is mediated by technology and the initiator and recipient are not co-located, it is harder for initiators to predict the receivers' current state (Fish, Kraut, Root, & Rice, 1992). Consequently, a large number of past projects focused on enabling spontaneous communication over a distance (see, for example, Dourish & Bly, 1992; Fish et al., 1992; Bly, Harrison, & Irwin, 1993; Adler & Henderson, 1994; S. E. Hudson & Smith, 1996). Further advances in communication technology, such as mobile phones, IM, and the growing availability of wireless networks, have lowered the barriers to initiating communication over a distance. These technological advances have been shifting communication from a place-to-place paradigm in which communication technology is tied to a locale, to a person-to-person paradigm in which communication technology is tied to an individual (Wellman, 2001). This, in turn, gives rise to a person's "reachability" and allows for an increase of spontaneous communication (of both work and social nature). However, unplanned spontaneous communication, whether technologically-mediated or not, does not come without a cost – specifically, the cost from interruptions.

2.1 The Disruptive Nature of Communication

The effect of interruptions on task performance, attitude, and wellbeing has been examined in a growing number of laboratory experiments (Gillie & Broadbent, 1989; Zijlstra, Roe, Leonova A.B., & Krediet, 1999; Bailey, Konstan, & Carlis, 2000; Czerwinski, Cutrell, & Horvitz, 2000b; Eyrolle & Cellier, 2000; Bailey, Konstan, & Carlis, 2001; Cutrell, Czerwinski, & Horvitz, 2001; McFarlane & Latorella, 2002; Monk, Boehm-Davis, & Trafton, 2002; Adamczyk & Bailey, 2004; Czerwinski, Horvitz, & Wilhite, 2004; Monk, 2004; Robertson, Prabhakararao, Burnett, Cook, Ruthruff, Beckwith, & Phalgune, 2004).

The disruptive effect of interruptions has been described to result from the introduction of new tasks on top of the ongoing activity, often unexpectedly. A person's limited processing and memory capacity results in conflicts between the current activity and the interrupting activity (Miyata & Norman, 1986). Experiments have consistently shown that performance on an ongoing primary task is hindered by interruptions (Gillie & Broadbent, 1989; Bailey et al., 2000; Czerwinski et al., 2000b; Eyrolle & Cellier, 2000; Bailey et al., 2001; Cutrell et al., 2001; McFarlane & Latorella, 2002; Monk et al., 2002; Adamczyk & Bailey, 2004; Czerwinski et al., 2004; Monk, 2004) and that interruptions may result in increased annoyance and anxiety (Bailey et al., 2001). An exception was reported by Zijlstra et al. (1999) where participants were able to develop strategies enabling them to deal effectively with interruptions, however, still having a negative effect on emotion and wellbeing.

The negative effect of interruptions has been shown to be sensitive to the type of the primary task (for example, Bailey et al., 2000), and to the type and length of the interrupting task and its similarity to the primary task, presumably because the two tasks are competing for similar attention resources (see, for example, Gillie & Broadbent, 1989). It has further been shown that the state of the primary task at the time of interruption had significant effects on subjects' performance on the secondary interruption and their ability to resume the primary task. Based on research showing the hierarchical structure of tasks into subtasks of different granularity (Zacks, Tversky, & Iyer, 2001), it has been shown that disruptions to a primary task are lower if interruptions arrive at task and sub-task boundaries (Adamczyk & Bailey, 2004; Iqbal, Adamczyk, Zheng, & Bailey, 2005). Monk et al. (2002) showed that the point of interruption in a primary task had significant effect on the time it took subjects to resume the task (with lowest resumption lag when interrupted just before beginning a new task stage). Similarly, subjects in an experiment by Cutrell et al. (2001) were interrupted while searching through a list. They found that interruptions harmed performance significantly more when occurring early in the search compared to interruptions towards the end of the search.

Outside the laboratory, the disruptive effect of interruptions and the cost to ongoing work has been observed through a series of field studies showing the high fragmentation of work as a result of interruptions (O'Conaill & Frohlich, 1995; Perlow, 1999; J. M. Hudson, Christensen, Kellogg, & Erickson, 2002; Czerwinski et al., 2004; Gonzalez & Mark, 2004; Mark, Gonzalez, & Harris, 2005). Participants in a field study on the multitasking of information workers demonstrated high levels of work fragmentation and interruptions (Gonzalez & Mark, 2004; Mark et al., 2005). Participants in this study were interrupted by others, on average, every four minutes throughout the work day (interestingly, when the participants were not interrupted by others, they were observed to interrupt themselves).

One of the main problems with such constant interruptions is the great difficulty to resume a task that has been interrupted. In a study on the nature of interruptions in the workplace, reported by O'Connaill and Frohlich (1995), two mobile professionals were observed. They report that recipients of interruptions returned to their original activity in only 55% of cases. Mark et al. (2005) found that participants in their study took over 25 minutes, on average to resume an interrupted task. They also found that, following an interruption, participants tended to engage in other activities before resuming the interrupted task (either through external or internal reminders). Iqbal and Horvitz (2007b) describe similar findings where participants in their study took more than 10 minutes, on average, to resume an interrupted task after tending to incoming IM, and about 16 minutes when tending to an incoming email. They also found that the time spent on the primary task before the interruption affected the likelihood that this task will be resumed, with shorter time on a task before an interruption corresponding with lower likelihood of resumption.

In fact, interruptions can be so disruptive to ongoing work that people will sometimes intentionally isolate themselves from communication. A study of research-managers and their handling of interruptions (J. M. Hudson et al., 2002) reported that some managers perceived interruptions to be such a problem that they would physically move away from their computer or even move away from their offices to avoid being interrupted. (In the particular case of Instant Messaging, relevant to this dissertation, I observed a number of managers who refused to use IM altogether for fear of being interrupted.)

Perlow (1999) observed how problematic reward structure in an organization led to disruptions at the individual level and in turn led to severe negative effects at the organizational level. In her study of engineers at a software company, she noted that engineers, whose work was delayed when interrupted with requests for help, would in turn interrupt when they needed help without regard for the other's work. This cycle, she observed, led to reduction in productivity, missed deadlines, and loss of money.

2.2 Combating Interruptions

A growing effort within the human-computer interaction community has focused on the identification and design of mechanisms for reducing the disruptive effects of interruptions. These include technologies for appropriate timing of interruptions, appropriate presentation of interruptions, and mechanisms for leveraging the social constructs within which communication-based interruptions are embedded.

2.2.1 Interruption Timing and Task Boundaries

As described above, the timing of an interruption, relative to the execution of a primary ongoing task, can make significant difference to the negative cost of the interruption (Cutrell et al., 2001; Zacks et al., 2001; Adamczyk & Bailey, 2004; Monk, 2004; Robertson et al., 2004; Iqbal et al., 2005).

McFarlane (2002) examined four methods of interruption delivery in human-computer interaction systems: *immediate*, in which the messages were delivered to the screen directly; *negotiated*, in which a notification flashed on-screen when a message arrived and the participant explicitly switched to the message to attend to it; *scheduled*, in which messages were delivered at preset intervals according to a schedule; and *mediated*, in which messages were delivered based on the participant's current workload in the primary task (the mediated delivery approach was extended by Dabbish and Kraut (2004) to interruptions originating from interpersonal communication and where the mediation was performed by the initiator of the communication).

The results of this study showed that performance on the primary task was significantly better in the *negotiated* condition; when subjects were able to defer interruptions until periods of low workload to engage in the interrupting task (similar results were presented by Robertson et al., 2004). The *negotiated* delivery method, however, also resulted in the worst timeliness in handling the interruptions. McFarlane notes that allowing users to negotiate the timing of an interruption (in the negotiated condition) could result in interruptions being delayed indefinitely. This problem, however, could be avoided through the use of a *bounded deferral* approach (Horvitz, Kadie, Peak, & Hovel, 2003). In this hybrid approach to timing of interruptions, notifications are deferred until the user transitions to a state of availability (Horvitz, Apacible, & Subramani, 2005a) or until the user enters a context that is defined by the sender of the interruption to be relevant (Jung, Persson, & Blom, 2005). After a pre-specified length of time, if the notification has not yet been delivered, it is presented immediately.

Other research suggested moments of physical transitions between activities as favorable for delivering interruptions (Ho & Intille, 2005). In this study, Ho and Intille compared subjects' receptivity to interruption when interruptions were delivered at activity transitions relative to those delivered at random times. Activity recognition for timing of interruptions was done using predictive models with accelerometers as source data (Bao & Intille, 2004), identifying transitions from sitting to walking, walking to sitting, sitting to standing, and standing to sitting. Participants in their study rated messages delivered at activity transitions significantly more favorably than messages delivered at random times.

Finally, findings showing that the cost of distractions is lower when people are interrupted at boundaries within a task hierarchy (Zacks et al., 2001; Adamczyk & Bailey, 2004; Iqbal et al., 2005) led to an important stream of work on identifying opportune moments for interruptions through automatic detection of task- and subtask-boundaries. Bailey, Adamczyk, Chang, and Chilson (2006), for example, developed a system that allows the monitoring of a user's progress through a task using pre-described task descriptions. Iqbal and Bailey (2007) presented predictive statistical models that learn to identify subtask boundaries based on labeled videos of a set of tasks performed in a laboratory settings. In a related experiment (Fogarty, Ko, Aung, Golden, Tang, & Hudson, 2005b), subjects performing programming tasks were interrupted with a contrived secondary task. Low-level events from the programming environment were then used to predict the latency of attending to the secondary interrupting task.

2.2.2 Interruption Presentation

One serious problem with many communication systems is the difficulty in distinguishing the importance and urgency of an interrupting communication from the notification of the communication (such as identical rings for incoming calls, the same flashing icon for all incoming instant messages, etc.). The inability to easily detect the potential importance, urgency, and relevance of an approaching interruption requires users to devote significant attention merely to choose whether or not to engage in the communication. Indeed, prior research has discussed and studied the importance of the design of notifications for reducing the negative effect of interruptions on performance and annoyance (Cutrell et al., 2001; Bartram, Ware, & Calvert, 2003; McCrickard, Catrambone, Chewar, & Stasko, 2003; McCrickard & Chewar, 2003; Gluck, Bunt, & McGrenere, 2007), proposing that the design of a notification and the attentional draw of the notification should correspond to attributes of the interruption, such as its importance and urgency.

In the vein of this prior work, the tool presented in Chapter 7 aims to provide differential notifications for incoming messages associated with differing response expectations.

2.2.3 Awareness and Contextual Information

In the special case of communication initiation, one possible way to reduce receiver interruptions is to include the initiator in the decision process by providing the initiators with contextual and awareness information about the receiver (see Milewski & Smith, 2000; Schmidt, Takaluoma, & Mäntyjärvi, 2000; Bellotti & Edwards, 2001; Pedersen, 2001; Tang, Yankelovich, Begole, Van Kleek, Li, & Bhalodia, 2001; Dabbish & Kraut, 2004; Avrahami, Gergle, Hudson, & Kiesler, 2007b).

The main benefit of this type of solution is that it re-distributes the interruption decision, removing some of the cognitive and social burden from receivers and placing it in the hands of initiators. For the initiator, a promising aspect of this type of solution is that it could leverage human judgment in determining whether the subject of conversation (often known only to the initiator) and the current social environment of the receiver yield an appropriate time for initiating communication.

Dabbish and Kraut (2004) found that awareness displays were able to significantly reduce the number of interruptions when participants in their study were provided with an awareness display of their partner's work-load. They found that an abstract display of the partner's workload resulted in the greatest reduction in interruptions. They also found that providing dyads with group identity resulted in initiators displaying greater sensitivity in timing their interruptions.

A study conducted by Avrahami et al. (2007b) examined the effectiveness of this type of solution by measuring the degree of agreement between receivers' desires and initiators' decisions. In their study, participants either played the role of Callers, deciding whether to interrupt a receiver with messages of varying importance and urgency, or the role of Receivers choosing whether they desire to be interrupted with each of the same messages. Their results showed that callers who were provided with contextual information about receivers made significantly more accurate decisions than those without it. Their results also suggest that different contextual information generate different kinds of improvements: more appropriate interruptions or better avoidance of inappropriate interruptions.

There are, however, a number of important issues concerning providing awareness and contextual information. First, the information provided to initiators may be insufficient to allow them to make appropriate decisions. Second, the information provided may be misinterpreted by initiators and used inappropriately. For example, Fogarty, Lai, and Christensen (2004b) hypothesized that the users of their MyVine system ignored the indications of availability provided by their system, using these indications, instead, to discover moments when their buddies were present. Similarly, previous research showed that different contextual information present in videos had significant correlation with biases in study participants' estimations of interruptibility of the videos' subjects (Avrahami, Fogarty, & Hudson, 2007a). In the case of awareness information that is the product of a statistical predictive model (of significant relevance to the work described in this dissertation) users, both initiators and receivers, may have difficulty forming an accurate mental model of the way in which predictions were arrived at (Tullio, Dey, Chalecki, & Fogarty, 2007). A third known problem associated with providing too detailed contextual information of a receiver's state is that initiators may spend so much time observing receiver's state in order to time their interruption that they will unnecessarily hurt their own performance (Dabbish & Kraut, 2004). Finally, providing detailed contextual information to initiators could compromise the privacy of the receivers. An in-situ study of user privacy preferences and patterns of sharing different types of context information with different social relations found that participants disclosed their context information generously, suggesting that the use of context information is feasible (Khalil & Connelly, 2006).

2.3 Systems and Modeling Approaches

As discussed above, prior research suggests that it should be possible to reduce the negative effect of interruptions through appropriate timing and appropriate presentation of the interruptions. But how can one identify moments that are good (or bad) for interruptions? In an attempt to answer this question and to assist in alleviating the negative impact of interruptions, a strong research drive has been growing in the past decade looking at the use of statistical methods to infer or predict a user's state and activity. These efforts have focused predominantly on inferring and predicting presence at a computer (for example, Horvitz, Koch, Kadie, & Jacobs, 2002; Begole, Tang, & Hill, 2003), attendance of meetings or events (Mynatt & Tullio, 2001; Horvitz et al., 2002; Tullio, Goecks, Mynatt, & Nguyen, 2002; Horvitz, Koch, Sarin, Apacible, & Subramani, 2005b), and a user's general cost of interruption or general level of interruptibility (for example, Horvitz, Jacobs, & Hovel, 1999; Horvitz et al., 2003; S. E. Hudson, Fogarty, Atkeson, Avrahami, Forlizzi, Kiesler, Lee, & Yang, 2003; Fogarty, Hudson, & Lai, 2004a; Iqbal & Bailey, 2006). While the vast majority of these works focused on office settings, a small number investigated interruptibility in the home (see, for example, Nagel, Hudson, & Abowd, 2004), in social settings (see Kern, Antifakos, Schiele, & Schwaninger, 2004), or, as in the work presented in this dissertation, left the setting unconstrained. Incorporating models of general interruptibility into systems has been met with varying degrees of success (see, for example, Begole, Matsakis, & Tang, 2004; Fogarty et al., 2004b).

With the Priorities system, Horvitz, Jacobs, and Hovel (1999) introduced a framework for the use of statistical models that infer a user's workload based on real-time sensing of the user's computer activity, calendar data, and other contextual information. The Priorities system showed the feasibility of automatically balancing the value gained from delivering an alert or communication (or cost of deferring) and the cost associated with interrupting the user with the alert. The value of the delivery of a message was estimated using textual analysis of a user's incoming email. Based on this cost-sensitive analysis, Priorities is then able to choose among different modes for delivering the alert (e.g., playing sounds that indicate the criticality of the message, bringing the client to the foreground, or even forwarding messages to a user's cell phone or pager). Presence forecasting – predicting the likelihood of a user returning to their computer within a period of time, given the user has already been away for some time – was added to a later version of Priorities (Horvitz et al., 2002). This version of the system included a component (SmartOOF) for sending custom-tailored messages back to senders telling when the receiver is predicted to next be available to read their messages. It also included a component (TimeWave) that posts indications of a user's unavailability on a shared calendar.

Bearing special relevance for the work described in this dissertation is the Coordinate system (Horvitz et al., 2002). Using data that is collected in an ambient fashion from multiple devices, Coordinate is able to perform forecasting of when a user will next transition to some state of interest, including the time until a user will return to their office, read their email, or be at a state of low interruption cost. The Coordinate system also includes models for estimating the likelihood that a user will attend a meeting or not and the cost of interrupting the user during the meeting (for additional work on attendance predictions see Mynatt & Tullio, 2001). Such learned models were later used, for example, in the Bayesphone system (Horvitz et al., 2005b), allowing a computationally-limited device to handle incoming calls intelligently. The Bayesphone work included an exploration of the use of real-time value of information in order to decide whether to collect training data from users.

Begole et al. presented visualizations and predictions of presence generated by examining records of minute-by-minute computer activity, the location of the activity, online calendar appointments, and e-mail activity (Begole, Tang, Smith, & Yankelovich, 2002). They attempted to predict the time until a user might resume activity and therefore become reachable for communication (note, *reachable*, not necessarily *available*). In follow up work, they examined different possible designs of such visualizations and predictions of presence (Begole et al., 2003). An important finding from their study relates to the inaccuracies in the model that related to changes in people's routines over time. This suggests that the relative weight of recent events should be considerably high.

A number of previous efforts were conducted to try and use statistical methods to model and infer a person's general state of interruptibility. That is, a measure of receptiveness to any form of interruption (be it a message that the computer's battery is fully charged or a colleague dropping by). To contrast, the work described in this dissertation focuses on communication-related interruptions for which the source of the interruption as well as the topic of interruption play significant roles (and the cost of deferral involves other people and ongoing relationships).

Our Wizard of Oz study examined the possibility of predicting general interruptibility from sensors (S. E. Hudson et al., 2003; Fogarty, Hudson, Atkeson, Avrahami, Forlizzi, Kiesler, Lee, & Yang, 2005a). Self-reports of interruptibility were collected from four office workers along with audio and video recordings. The self-reports were collected through an experience sampling method[*] (Csikszentmihalyi, Larson, & Prescott, 1977) by interrupting participants at random times and asking them to report their interruptibility (to an unspecified interruption) on a 5-point scale. The audio and video recordings were then hand-coded to simulate a wide range of possible sensors (for example, the state of the door, the use of the telephone, and the presence of guests). Statistical models, created from these simulated sensors to predict the self-reported interruptibility, were able to predict with high accuracy

[*] Experience Sampling Method, or *ESM*, refers to a data collection method in which participants, functioning within their natural settings, respond to repeated probes presented over time.

times of high non-interruptibility. These models suggested that sensors that can detect whether someone in the office was talking were useful for identifying times that participants were not interruptible. In a follow-up study, Fogarty, Hudson, and Lai (2004a) deployed actual sensors to examine models of self-reported interruptibility of a broader set of office workers (managers, researchers, and interns). One of their interesting findings was that a sensor that detects whether someone in the office was talking was very useful when participants had private offices, however, not as useful when participants occupied a shared space.

Another work examining the ability to model a person's general cost of interruptions is the Interruption Workbench (Horvitz & Apacible, 2003). In this work, three participants reviewed their own audio and video recordings to provide labels of their cost of interruptions on a 3-point scale at different times. These labels were used to train predictive models based on participants' computer activity, visual and acoustical analyses of the recordings, and calendar data. One interesting aspect of the Interruption Workbench is the forecasting of the time until a user will be at one of the three levels of costs of interruption. Such forecasting, can allow other people, as well as applications, to make complex decisions regarding the deferral of interruptions. The BusyBody system (Horvitz, Koch, & Apacible, 2004) predicts the cost of interruption (for some general interruption) on a two point scale (Busy vs. Not Busy) based on self-reports gathered using an experience sampling method. BusyBody uses dynamic Bayesian networks to analyze the relationship between the collected self-reports and sensors related to the desktop event stream, time of day, day of week, electronic calendar events, a microphone-based conversation detection system, and WiFi-based location estimates.

In contrast to the prior research described above, the work presented in this dissertation looks to combine three important aspects that have often been overlooked by each of the individual pieces of work. My work focuses on the collection and use of naturally occurring interaction (field-data). These collected data are used for the construction of statistical predictive models, but also for an examination of the underlying (naturally occurring) interaction, potentially providing insights into the accuracy achieved by the predictive models. Finally, my work on responsiveness to communication allows for the construction of predictive models based purely on explicit, observable measures.

2.4 Between Asynchronous and Synchronous Communication

In the work presented in this document, I have focused on investigating and enhancing Instant Messaging communication.

Interpersonal communication through Instant Messaging, or *IM*, is gaining increasing popularity in the work place and elsewhere. A report from 2005 estimated that 12 billion instant messages are sent each day. Of those, nearly one billion messages are exchanged by 28 million business users (Mahowald, 2005). IM programs, or *clients*, facilitate one-on-one communication between a user and their list of contacts, commonly referred to as *buddies*, by allowing them to easily send and receive short textual messages ("*instant messages*"). Figure 2.1 shows a screenshot of a computer desktop displaying an IM client with a buddy-list, and an IM message window.

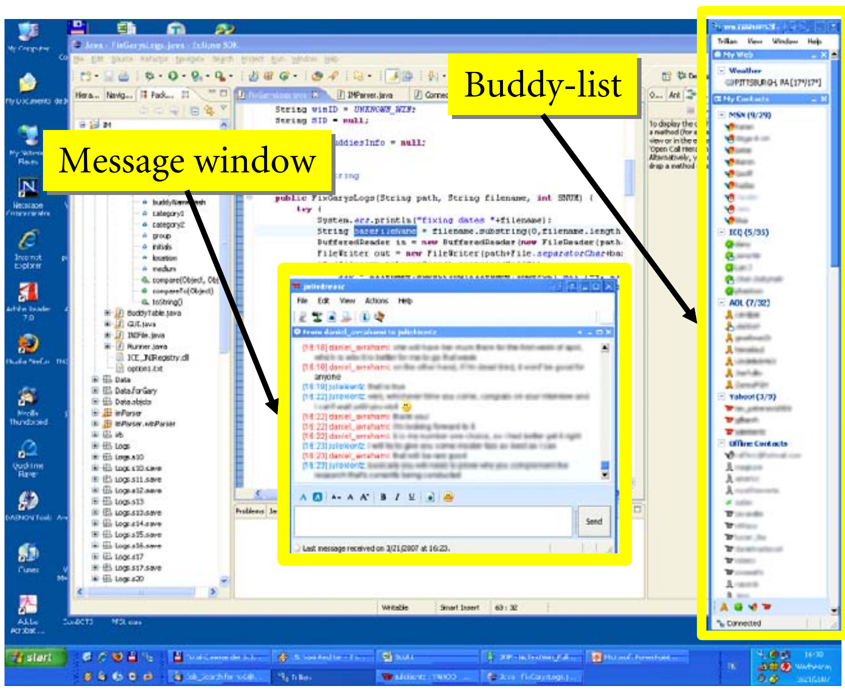


Figure 2.1 An IM buddy-list (on the right) and an IM message-window (center) on a computer desktop.

Despite its popularity, IM suffers from a number of shortcomings. Specifically, the ease of initiating communication, combined with limited awareness of receivers' state, result, as illustrated above, in messages often arriving at inconvenient or disruptive times for the receiver.

Instant messaging was introduced in 1996 by the Israeli startup Mirabilis with their ICQ messaging service ("Mirabilis Inc."). In its early days, IM gained its widest use supporting social communication, primarily between teenagers. As reported by Grinter and Palen (2002), teens used IM primarily for socializing and planning social events, but also for coordinating schoolwork. When IM was introduced into the workplace, it was thus often met with resistance, being perceived as a medium suitable primarily for social communication (Slatalla, 1999; Herbsleb, Atkins, Boyer, Handel, & Finholt, 2002). More recently, however, organizations are recognizing the value of IM and its benefits as a lightweight communication medium. Research showed that IM communication in the workplace has many uses and benefits in complementing other communication mediums. These uses range from quick questions and clarifications, coordination and scheduling, to discussions of complex work (Bradner, Kellogg, & Erickson, 1999; Nardi, Whittaker, & Bradner, 2000; Handel & Herbsleb, 2002; Herbsleb et al., 2002; Isaacs, Walendowski, Whittaker, Schiano, & Kamm, 2002).

Figure 2.2 presents a single real IM session from the data collected in this work. This session was exchanged between one of my participants and one of their buddies, a co-worker. This session illustrates the lightweight nature of IM communication. In fewer than two minutes, and using no more than 12 messages, both participant and buddy were able to exchange brief greetings (messages# 1 and 3), coordinate a simple task (messages# 2,4,6,7), and apologize (message# 11) for a typing error made more than 30 seconds earlier (message# 8). This session also illustrates the use of abbreviations, loose grammar and minimal punctuation, prevalent in IM (Nardi et al., 2000; Voida, Newstetter, & Mynatt, 2002).

#	Time		Message Text
1	17:42:45	B:	Hey [*Participant's name*]
2	17:42:56	B:	what time does your group get in the AM?
3	17:42:57	P:	hey
4	17:43:01	P:	usually around 10
5	17:43:25	B:	ok
6	17:43:38	B:	i want to start circulating the card in the AM
7	17:43:58	P:	ok, good idea
8*	17:44:02	P:	that's for coordinating this
9	17:44:13	B:	no problem
10	17:44:27	P:	thanks :-)
11	17:44:35	P:	sorry bout the typo
12	17:44:38	B:	is ok

* The participant meant to write "thanks" and not "that's"

Figure 2.2 A single IM session between one of the participants (P) and a buddy who is their co-worker (B).

A number of benefits of IM have contributed to its increasing popularity. While IM, in its underlying architecture is asynchronous, its lightweight nature allows conversation to range from rapid exchanges of messages, to hours and even days passing between messages in the same conversation (Nardi et al., 2000). Thus IM is often described as a "near-synchronous" communication medium, positioned somewhere between synchronous communication channels (such as phone or face-to-face) and asynchronous communication channels (such as email, newsgroups, and online forums). Voida et al. (2002) attribute a number of interesting behaviors of IM users, such as their need to acknowledge typing errors, to the tension between the near-synchronous yet still asynchronous and persistent nature of IM dialog.

Since IM is inherently asynchronous, users can choose when or whether to respond to an incoming message. As noted by Nardi et al. (2000), the limited presence information in IM provides users with *plausible deniability* when they elect to ignore or postpone responding to a message (that is, users can easily claim to not have seen a message, or claim to not have been present). IM is thus often regarded as less disruptive than other synchronous communication channels. In fact, IM is sometimes used for communication even between users who share the same physical workspace in an attempt not to disrupt one another's work. This asynchrony, however, means that messages often arrive when a user is engaged in other tasks. Indeed, research shows that users often multitask when using IM (Nardi et al., 2000; Grinter & Palen, 2002; Isaacs et al., 2002). Particularly in the work place, messages may thus arrive when a user is engaged in important and potentially urgent work. Staying on task and not responding may come at a cost to the initiator, who may need some information from the receiver. The receiver herself may incur a social cost from being portrayed as unresponsive. Engaging in conversation, on the other hand, will often come at a cost to the receiver's ongoing work (Voida et al., 2002).

In previous studies of the effect of interruptions, Gillie and Broadbent (1989) showed that even a very short interruption can be disruptive, while Cutrell et al. (2001) showed that even an ignored interruption can have a negative effect. Czerwinski et al. (Czerwinski, Cutrell, & Horvitz, 2000a) showed the relationship between the effect of an interrupting incoming message and the user's ongoing task and its relationship with the user's position in the task. Taken together these results indicate that an incoming instant message, even if ignored, can have a negative effect on the user's ongoing work.

One of the most important features of IM clients is the ability to provide some awareness of *presence.* IM clients typically provide this information by indicating whether a user is online and whether the user is currently active or idle (often referred to as the user's "Online Status"). Most IM clients also allow users to set additional indicators to signal whether they are busy or away from the computer. Those, however, are often insufficient as they require users to remember to set and reset them (Milewski & Smith, 2000). Begole et al. presented a system that was able to predict a person's presence based on observed patterns (Begole et al., 2002).

Knowing whether a person is *present*, however, does not necessarily provide an indication of whether or not that person is *available* for communication (Begole et al., 2004; Fogarty et al., 2004b). A user who is not present (typically indicated as 'offline' or 'idle') is indeed not available for communication. On the other hand, a user engaged in an important task will be indicated by an IM client as present (unless they remembered to manually set their status to 'Busy') but may in fact be unavailable for communication.

Since the content or topic of an incoming communication is typically unknown to the user before it arrives, users generally have to attend to all messages. While the tool presented in Chapter 7 increases alerts to some messages based on their content, it does not prevent default alerts from taking place. As a result, users may even elect to turn their IM client off when they are busy, refusing incoming messages altogether (Nardi et al., 2000; Hafner, 2003). As Isaacs et al. (2002) note, however, most IM conversations held in the workplace are work-related. This makes closing the IM client a less desirable strategy. Similar to the use of Caller ID in phones, a user can typically also see who the sender of the message is before attending to the message. However, even this brief interruption can, in and of itself, be disruptive (see, for example, Gillie & Broadbent, 1989). Results from Dabbish and Kraut (2004) and Avrahami et al (2007b) suggest that, given information about the receiver, senders would be able, and willing, to time their messages to accommodate for the receiver's state.

In this document I describe my research aimed at enhancing interpersonal communication over Instant Messaging by providing a better understanding of factors affecting communication in context, and through the creation of predictive statistical models trained using *naturally occurring human behavior*.

CHAPTER THREE

Predicting Responsiveness to IM*

3.1 Motivation

Incoming instant messages join an ever growing number of interruptions a person is exposed to. Those include interruptions external to the computer, such as telephone calls or people stopping by to ask a question, as well as interruptions from various computer applications, including alerts of incoming email, calendar notifications, or notifications of new items from RSS feeds.

Unlike face-to-face communication, users of IM cannot easily detect whether a buddy is available for communication or not. The inability to detect a buddy's state can often result in communication breakdowns with negative effects on both communication partners. For the receiver, communication at the wrong time might be disruptive to their ongoing work. If, on

---

* The work presented in this chapter was originally published in Avrahami, D., & Hudson, S. E. (2006). Responsiveness in Instant Messaging: Predictive Models Supporting Inter-Personal Communication. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2006)*, pp. 731-740. ACM Press.

the other hand, receivers simply decide to ignore communication, the initiator's productivity could suffer if they are left waiting for a piece of information.

The ability to predict responsiveness, in IM and other mediums could provide a number of benefits to communication partners. In the work presented in Chapter 7, I created an augmentation to an IM client that allowed users to project different "responsiveness images" in IM (Avrahami & Hudson, 2004). As users of Computer-Mediated-Communication (CMC) typically have limited awareness of the state and context of the remote conversation partner, slow or no responsiveness in these situations can be easily misinterpreted. Herbslab et al. (2002) found that, in accordance with the actor-observer effect (Jones & Nisbett, 1971), users will often attribute lack of responsiveness to internal causes such as personality traits of the conversation partner ("person attribution") rather than to external causes ("situation attribution").

If, however, we were able to accurately predict whether a user was likely to respond to a message within a certain period of time, then some of the breakdowns (of both interruptions and attribution) could be prevented. For example, models could be used to automatically provide different "traditional" online-status indicators to different buddies. Alternatively, models can be used to increase the salience of incoming messages that may deserve immediate attention if responsiveness is predicted to be low. Models could also be used by a system that will show a list of potentially responsive buddies to users who are looking for help or support, while hiding others.

In previous work (S. E. Hudson et al., 2003) we have demonstrated the ability to create statistical models that predicted, with relatively high accuracy, time periods reported by participants as highly non-interruptible. The models presented in that work predicted interruptibility for a general-non specific interruption (including, for example, a notification from an operating system that the computer battery is fully charged). Hovitz et al., for example, presented statistical models that were able to predict whether a user is “Busy” or “Not Busy” with accuracy as high as 87% (Horvitz et al., 2004).

3.1.1 From Availability to Responsiveness

Availability for inter-personal communication is a concept not easy to define. Many factors can contribute to a person’s availability: their current mental task, the proximity to the next breakpoint, the identity of the conversation partner, established organizational norms and culture, and so on.

Unfortunately, getting at a person’s “true” availability is near impossible. Furthermore, a person’s *stated* availability, how available they claim to be, may not match their *demonstrated availability* – their actual responsiveness to communication. For example, a person may be busy and *state* that they are unavailable for communication, while organizational norms coerce that same person to respond to incoming communication (Ghosh, Yates, &

Orlikowski, 2004; Rennecker & Godwin, 2005), thus *demonstrating availability*. While stated availability is of great interest to us and others, I have decided to focus my initial efforts on predictions of demonstrated availability, more specifically, on the ability to predict *responsiveness* to incoming communication. Tyler and Tang (2003) investigated responsiveness to email through interviews and observations. They found that users modified their own levels of responsiveness in order to project different "responsiveness images". For example, they used responsiveness to provide others with an indication of both availability, and also of their perception of the importance of a message. It is my hope that this work will allow us to further understand the relationship between responsiveness, demonstrated availability, and finally availability for communication overall.

3.1.2 Behavior as Ground Truth

In order to create a predictive model using machine learning techniques referred to as *supervised learning*, one must first gather data along with *labels* that represent *ground truth* about the data. (Other machine learning techniques, referred to as *unsupervised learning*, that do not use labeled data also exist, but are often less useful for HCI purposes). For example, a set of email messages along with labels provided by a user, indicating messages as either 'spam' or 'legitimate', can be used to train a model to identify spam email messages.

Previous related work, including (Horvitz & Apacible, 2003; S. E. Hudson et al., 2003; Fogarty et al., 2004a; Horvitz et al., 2004; Nagel et al., 2004), collected naturally occurring behavior as data, using participants' self reports as the labels of ground truth. Other work used the behavior of subjects participating in a lab experiment to create their predictive models (see, for example, Fogarty et al., 2005b; Iqbal & Bailey, 2006, 2007). For example, in the work presented in (S. E. Hudson et al., 2003; Fogarty et al., 2004a; Fogarty et al., 2005a) and used by Begole et al. (2004) for their models, labeled data were gathered by asking participants, at different intervals, to provide self-reports of their interruptibility on a scale of 1-5. Similarly, Horvitz et al. asked participants to indicate at random times whether they were busy or not busy (Horvitz et al., 2004). Horvitz and Apacible (2003) asked participants to observe video recordings of their day and retrospectively assign a monetary value to a hypothetical interruption. Nagel et al. had participants fill out a short survey on a PDA at random intervals (Nagel et al., 2004). Finally, Iqbal and Bailey (2007) employed observers to review videos in order to identify breakpoints in user interaction.

One of the main drawbacks of using self-reports as measures of ground truth, faced in previous work, is that they are very demanding from the participant's point of view and make it hard to collect large amounts of data. Responding to a voice-prompt (as in S. E. Hudson et al., 2003) or to a survey on a PDA (as in Nagel et al., 2004) or sitting for a long period of time to label past events (as in Horvitz & Apacible, 2003) can be socially and attentionally costly, and quite time consuming. Another problem with self-reports is that they reflect individuals' subjective interpretation of what is asked of them, an interpretation that can vary from individual to individual (for an interesting discussion of the strengths and weaknesses of the experience sampling method, see Scollon, Kim-Prieto, & Diener, 2003).

I should note that recent work has started addressing the problem of the high cost of labeling by investigating the possibility taking into account the state of the human labeler and the potential value gained from the additional label in order to determine the whether to request the user for the label (Kapoor & Horvitz, 2007; Kapoor, Horvitz, & Basu, 2007). For example, they present an adapted version of the BusyBody system (Horvitz et al., 2004) that will probe the user for a label only if the predicted value gained by the probe is high (Kapoor & Horvitz, 2007).

Models generated based on data collected in laboratory studies (such as Fogarty et al., 2005b; Iqbal & Bailey, 2006, 2007), provide valuable insights into fine grain factors that may be used to predict availability, interruptibility, or their variants (e.g., "Cost Of Interruption"). In previous work, for example, Iqbal, Adamczyk, Zheng, and Bailey (2005) found a relationship between the point of delivering an interruption during a task structure (reflected in its GOMS model structure), and the time needed to resume that task. Iqbal and Bailey (2006) then used this structure to create a classifier that predicts, albeit with very low accuracy, the cost of delivering an interruption at different points in the task. A main drawback of lab studies, however, is that their focus is often limited, in order to maintain experimental control (for example, using attention demanding but non-realistic interruptions), and their models are thus difficult to generalize.

In contrast with the work mentioned above (but similar to Begole et al., 2002; Horvitz et al., 2002), the work presented in this chapter describes the creation of predictive statistical models trained using *naturally occurring human behavior*. This is possible since responsiveness is a readily observable behavior. One added benefit of using naturally occurring behavior as the source for learning is that a model deployed as part of a system would be able to continuously observe user behavior to train and improve its performance without requiring any intervention from the user.

3.2 Data Collection Method

Data collection for this work was done using a background process implemented as a custom plug-in module for Trillian Pro, a commercial IM client developed by Cerulean Studios ("Cerulean Studios - Trillian Pro"), which runs on Windows operating system. I chose to use Trillian Pro as it supports the development of dedicated plug-ins through a Software Development Kit (SDK) giving access to most of the client's functionality.

Like a number of other IM clients, Trillian allows a user to connect to any of the major IM services (ICQ, AOL, MSN, Yahoo!, and IRC) from within one application.
Trillian Pro is further capable of communication with other IM services, including Jabber and Lotus Sametime. Using Trillian Pro thus allows me to recruit participants without

concern for the specific IM service that they use. (In fact, 10 of the 19 participants in my current data set used Trillian to communicate with buddies over two or more IM services during their participation, and using Trillian Pro allowed me to observe their interactions over all channels.)

I decided to use a commercial client rather than develop a client on my own because it provides functionality beyond the simple exchange of text messages. For example, it allows file sharing, audio and video chats, and sending images. Allowing participants this range of capabilities reduces the likelihood of participants using other IM clients that support these features during the course of their participation in the study.

To capture instant messaging events, as well as desktop events, a copy of Trillian Pro was purchased for each participant and then instrumented with the data recording custom plugin. The plugin is written in C and implemented as a Dynamically-Linked-Library (DLL) that is run from inside Trillian Pro. The plugin automatically starts and stops whenever Trillian Pro is started or stopped by the participant.

The following events are recorded by the plugin:

IM events:

- Message sent or received
- Trillian start or stop
- Message window open or close

- Starting to type a message
- Status changes (online, away, occupied, etc.) of both participants' and buddies'.
- Indicator for incoming message is blinking (if this setting is used)

Desktop events:

- Key press (does NOT include which key was pressed)
- Mouse button click / double-click
- Mouse move
- Window created (including window title and size of window)
- Window minimized (including window title)
- Window in focus (including window title and size of window)
- Window closed

These events, along with the time at which they occurred are saved into log files. These log files are compressed by the plugin "on-the-fly", encrypted, and then stored locally on participants' machines.

The compressed log files, along with the coding, were collected from participants' computers at the end of their participation and instructions were given to them for removing the plug-in.

Participants were instructed to use Trillian Pro for all their IM interactions for a period of at least four weeks.

3.2.1 Privacy of Data

The data collection mechanism includes a number of measures intended to preserve, as much as possible, the privacy of participants and their buddies. Unless participants provide specific permission, the text of messages is not recorded and messages are masked in the following fashion: Each alpha character is substituted with the character 'A' and every digit is substituted with the character 'D'. Punctuation is left intact. For example, the message "my PIN is 1234 :-)" is recorded as "AA AAA AA DDDD :-)". A simple mechanism for masking individual sessions is also provided to participants who allowed the recording of the text of messages; if a participant or buddy enters the string "/mm" in a message, that message and messages that follow (until the window is closed) are masked. (This mechanism was used occasionally by the participants and their buddies.)

When a participant opens a message window to a buddy for the first time (and that buddy is online), the following alert is sent to the buddy notifying them of the participation in the study: "This user is participating in a study and her/his IM is being logged. The text of messages is NOT recorded." Buddies of participants who had provided the additional permission to record the text of messages are notified with a different alert message that instructs them of a simple mechanism that allows them to temporarily mask messages ("This user is participating in a study and her/his IM is being recorded. You can prevent a message from being recorded by typing \mm anywhere in the message").

Finally, for determining that two events are associated with the same buddy, I create a unique ID for each buddy (using an MD5 cryptographic hash) and store the ID of the buddy instead of the buddy-name itself.

3.3 Participants

Using the data collection mechanism described above, I collected a total of approximately 6,600 hours of recorded data, observing over 125,000 incoming and outgoing instant messages from 19 participants in three phases.

The participants included eight Masters students at our department, eight employees of a large industrial research laboratory, and three employees at a local high-tech startup. Of the researchers, six were full time employees (three first-line managers and three full-time researchers) and two were summer interns. All participants used IM in the course of their everyday work. I will refer to the first eight participants as the *Students* group, the six full-time employees as *Researchers*, the two interns as *Interns*, and the startup employees as the *Startup* group (the data of the Startup group was used only in the work presented in Chapters 5 and 7).

The first data collection phase, which started in May 2005, included the data of the Students group. During their participation, each of these participants was engaged in a number of group projects as part of their studies. Of the Students, six were female and two male, with an average age of 24.5 (SD=2.39, Min=22, Max=29). Six of these participants ran the recording software on their personal laptops. One participant, who used a laptop at school and a desktop computer at home, ran the recording software on both machines. The eighth participant ran the recording software on his account on a shared desktop computer in the Masters students' lab. During their participation, each of these participants was engaged in a number of group projects as part of their studies.

In the second phase, which started in July 2005, I collected data of the Researchers and Interns groups. The average age of the six Researchers was 40.33 (SD=4.97, Min=34, Max=49) with three female and three male. One female and one male, the Interns group had an average age of 34.5 (SD=3.54, Min=32, Max=37). All participants in phase 2 ran the recording software on their work laptops. For confidentiality reasons, I did not record the text of messages from any of the participants in the Researchers or Interns groups.

In the third phase, which took place during the second half of 2006, I collected the data of the Startup group. This group included two females and one male. (A fourth participant from this group requested to withdraw from the study and his data was discarded.) The average age of the participants in this group was 32 (SD=7.5, Min=25, Max=40). All three participants allowed me to record the text of their messages.

The majority of participants were new to Trillian Pro but were able to automatically import the list of all their buddies into Trillian Pro. None of the participants had any difficulty making the transition to using Trillian Pro (and a few still use it now after the end of their participation), although some assistance was required with customization of specific options to match the preferences that individual users were accustomed to. All of the participants were required to run the recording software for a period of at least 4 weeks. A small number of the participants voluntarily continued their participation for longer time-periods.

3.4 Data Overview

Using Trillian Pro as the client on which the data collection was based resulted in the successful recording of a very high volume of IM events. (A small number of data files were unusable due to corruption in the on-the-fly compression, often as a result of participants' laptops running out of power.) Table 3.1 provides a summary of data collected in all phases. I collected a total of approximately 6500 hours of recorded data, observing over 125,000 incoming and outgoing instant messages. 73,906 messages from participants of phase 1 spread over 3,839 recorded hours, 17,633 messages in phase 2 from 1355 hours of recordings, and 34670 messages from participants of phase 3 spread over 1376 hours of recordings.

Two of the participants in the Researchers group recorded significantly fewer messages in their logs (96 and 350 messages). However, I did not remove their data from my models and analyses.

Table 3.1 Overview of the data collected from each participation group.

Participation Group	N	Avg age	Total hours recorded*	Avg hours recorded per participant per day	Total active buddies	Avg active buddies per participant	Total sessions	Total msgs	Avg msg per recorded hour
Researchers	6	40.3	982.5	6.4	130	21.7	845	7290	7.4
Interns	2	34.5	373.0	5.6	61	30.5	757	10343	27.7
Students	8	24.5	3839.8	9.4	244	30.5	2903	73906	19.2
Startup**	3	32.0	1376.2	6.9	56	18.7	2871	34670	25.2
Overall	19	31.7	6571.5	7.9	491	25.8	7376	126209	19.2

* These numbers do not include a small amount of data lost to corrupted log files.
** The data of this group was used only in the work presented in Chapters 3, 4, 5 and 7.

To accommodate the fact that data were recorded only when Trillian was running, I provide separate fields in Table 3.1 indicating the amount of time recorded, as well as the total participation time (calculated for each participant from the start time of their first log file, until the end time of their final log). Since participants in the second and third phase recorded activity primarily during business days, their participation time is multiplied by 5/7. The number of recorded hours per day did not vary significantly between groups (p=.23, N.S.).

Overall, message exchanges between the participants and their buddies demonstrated patterns of bursts of rapid exchanges followed by periods of inactivity. Figure 3.1 shows the delay between each 500 consecutive messages between one of our participants and one of their buddies. This pattern is similar to the pattern of email exchanges discussed in prior research (see, for example, Barabási, 2005; Kalman & Rafaeli, 2005) .

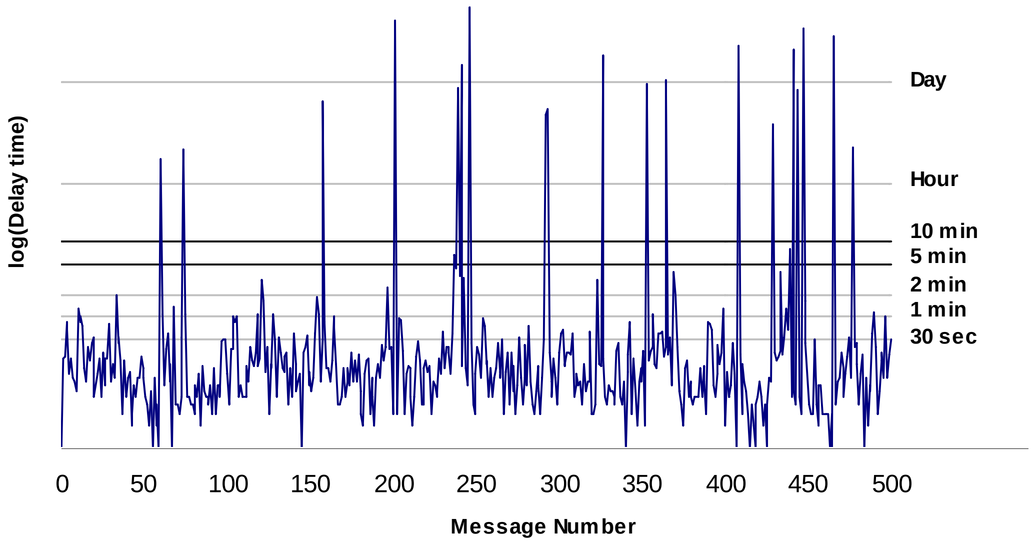


Figure 3.1 Delay (log sec) between 500 consecutive messages exchanged between one participant and one of their buddies.

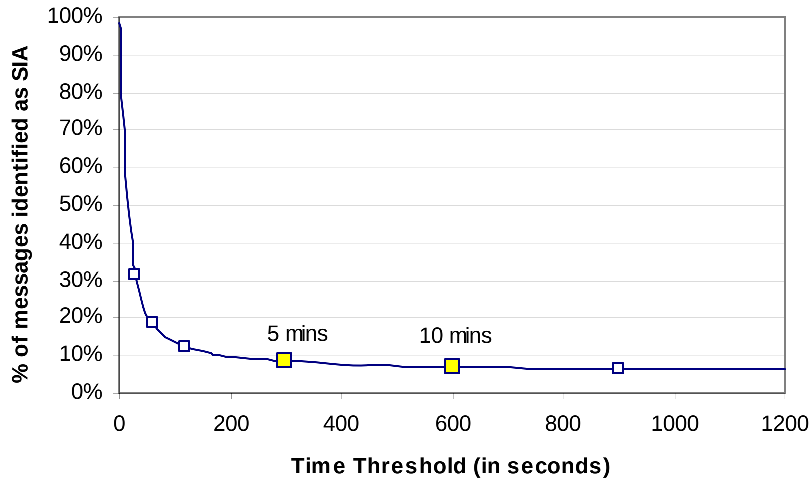
In my data set, 90% of messages are responded to within 5 minutes (in fact, 50% of the messages in my data are responded to within 18 seconds). This means that a system that always predicts that a user will respond to any incoming message within 5 minutes will be correct 90% of the time. However, the majority of messages occur as part of a rapid exchange of messages – what I call an *IM session*. Once a session has been established, responsiveness is likely to be high and can be explicitly negotiated between parties if needed (for example, one could explicitly declare reduced responsiveness by sending a message saying that a visitor has entered the room). Consequently, predicting responsiveness to an incoming instant message is interesting primarily for messages that can be defined as initiating a new session, rather than those inside a session proper.

3.5 Defining IM Sessions

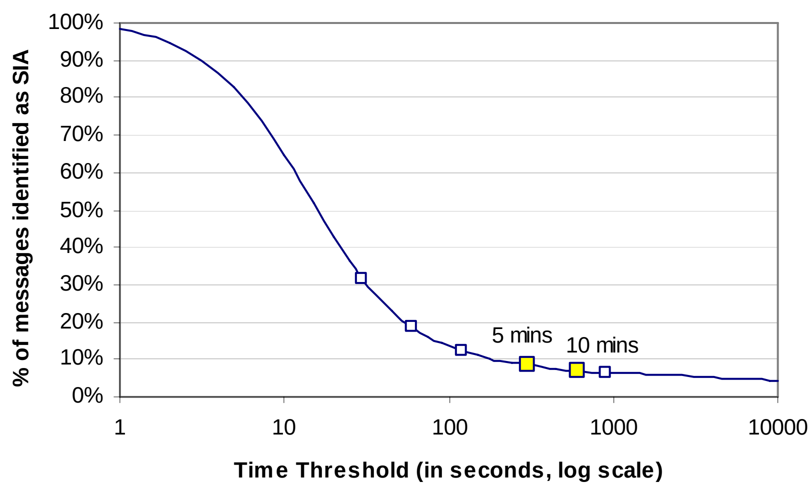
For the predictions and analysis of Responsiveness, as well as for the analysis and predictions of the effect of interpersonal relationship on communication, presented in the next section, I define an *IM session* to be a set of instant messages that are exchanged within certain time proximity of one another. That is, two consecutive instant messages are categorized as belonging to the same IM session if they were exchanged between a participant and their buddy within a certain time of one another. Unlike a conversation, a session is not determined by the content of its messages. Indeed, a single conversation may extend over multiple sessions, while a particular session may contain several conversations. Once a session has started, users will often explicitly state their forthcoming responsiveness (for example, by declaring themselves busy or notifying their buddy that they must leave for a short while). However, of particular interest would be the successful prediction of responsiveness to incoming communication *before* a session has started. Such prediction could help users decide whether or not to attempt to initiate a session with a buddy.

3.6 Session Initiation Attempts (SIA)

For the purpose of predicting responsiveness before a session begins, I define the concept of a Session Initiation Attempt. An incoming message from a buddy is identified as a *Session Initiation Attempt* (SIA) if the time that has passed since the participant sent a message to that same buddy is greater than some threshold.



(a)



(b)

Figure 3.2 Percent of messages identified as a Session Initiation Attempt (SIA) Time-threshold (in linear scale and logarithmic scale). Time thresholds are used to determine that a message belongs to a new session. The 5 and 10 minutes threshold used in this work are highlighted. Thresholds of 30 seconds, 1, 2, and 15 minutes are also highlighted for comparison.

The choice of the appropriate threshold to use in order to identify messages as SIA is not trivial. Figure 3.2 shows the percent of messages that are identified as Session Initiation

Attempts based on the time threshold used to determine whether one session ended and another is starting.

In the work presented in this chapter I have decided to use two thresholds (highlighted in Figure 3.2): a 5-minutes threshold (SIA-5), similar to the threshold used by Isaacs et al. (2002), and a more conservative 10-minutes threshold (SIA-10). Note that any message identified as a SIA-10 is necessarily also identified as a SIA-5. Of the 45,468 incoming messages in the data of the Researchers, Intern, and Students, 3,805 were identified as SIA-5 and 3,161 as SIA-10 (both session thresholds are indicated in Figure 3.1). 72% of messages in SIA-5 and 71% of messages in SIA-10 were responded to within 5 minutes, compared to 90% of the full set of messages. The median response time for messages in SIA-5 and in SIA-10 was 37 seconds, compared to the median of 17 seconds for all messages.

Table 3.2 Partial list of generated features.

Day of week
Hour
Is the Message-Window open
Buddy status (e.g., "Away")
Buddy status duration
Time since msg to buddy
Time since msg from another buddy
Any msg from other in the last 5 minutes
log(time since msg with any buddy)
Is an SIA-5

(a) IM features

App. in focus
App. in focus duration
Previous app. in focus
Previous app. in focus duration
Most used app. in past *m* minutes
Duration for most used app. in past *m* minutes
Number of app. switches in past *m* minutes
Amount of keyboard activity in past *m* minutes
Amount of mouse activity in past *m* minutes
Mouse movement distance in past *m* minutes

(b) Desktop features

3.7 Features and Classes

3.7.1 Features

The raw user-data was first processed to produce, for every incoming or outgoing message, a set of 82 features describing IM and desktop states and a set of classes that the models should learn. Table 3.2a shows a partial list of the IM features associated with every message. I adapted the desktop features from features used in (Fogarty et al., 2004a; Horvitz et al., 2004). Those include the amount of user activity and the most-used application, in the 0.5, 1, 2, 5, and 10 minutes time intervals that precede the message arrival time. I associated applications with a general set of application types (including for example, email, WWW, design-tool, etc.). Table 3.2b shows a partial list of the desktop features associated with every message.

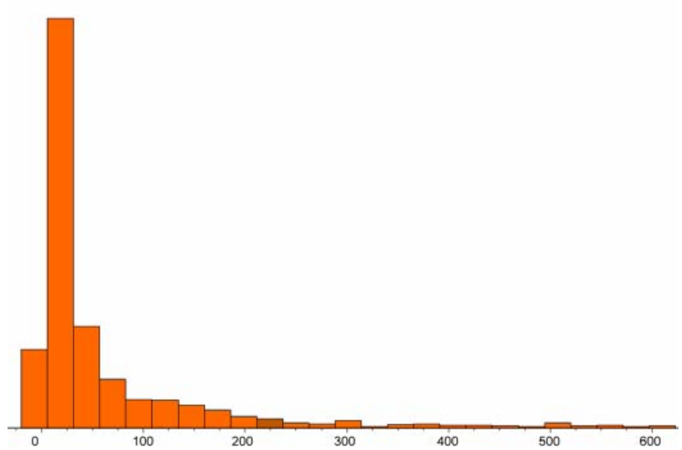


Figure 3.3 Histogram of "Seconds until Response" for incoming SIA-5 set with a cut-off at 10 minutes.

3.7.2 What is Predicted? (Classes)

My base measure of responsiveness, "Seconds until Response", was computed, for every incoming message from a buddy, by noting the time it took until a message was sent to the same buddy. A histogram of "Seconds until Response" for incoming SIA-5 messages is presented in Figure 3.3. From this base measure I then created five binary classification labels by indicating, for every message, whether or not it was responded to within each of the following five time periods: 30 seconds, 1, 2, 5, and 10 minutes. (Note that, as indicated in the previous section, less than half the SIA messages were responded to within 30 seconds, while more than half were responded to within the 1, 2, 5, and 10 minutes time periods).

I was now ready to train models to predict each of these binary classifications using the generated features.

3.8 Models Performance

This section presents the performance of statistical models of responsiveness to instant messaging, more specifically to Session Initiation Attempts over each of the classes described above. The models presented were generated using a J4.8 Decision-Tree classifier (an implementation of the C4.5 rev. 8 algorithm) using the Weka machine-learning tool-kit (Witten & Frank, 1999). Other classification techniques were also explored but generated models with lower accuracy. For the decision-tree models I used a wrapper-based feature selection technique (Kohavi & John, 1997). This technique selects a subset of the available features by incrementally adding features to the model and testing the model performance until no added feature improves the performance of the model. Each of the models in the process is evaluated using a 10-fold cross-validation technique. That is, each model is created over 10 trials, with each trial using 90% of the data to train, and the remaining 10% to test the model's performance. The overall model accuracy is then presented as the combined accuracy over these 10 trials. Finally, a boosting process took place using the AdaBoost algorithm (Freund & Schapire, 1996).

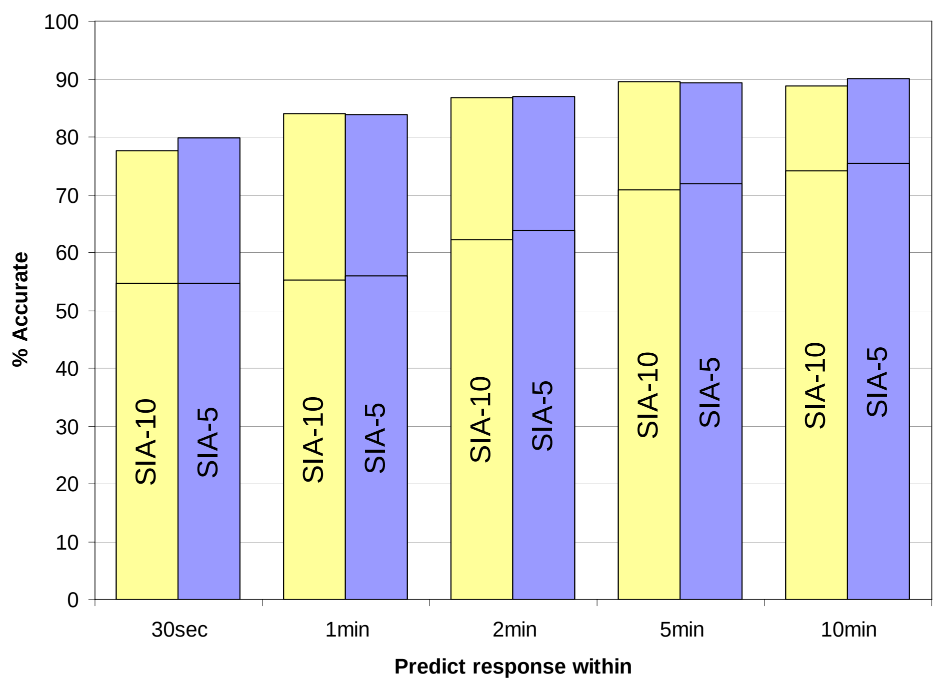


Figure 3.4 Accuracy of models predicting response to Session Initiation Attempts (SIA-5 and SIA-10) within 30 seconds, 1, 2, 5, and 10 minutes. Baseline prior probability is shown with the black lines.

The performance of ten models created for both SIA thresholds and predicting responses within 0.5, 1, 2, 5, and 10 minutes, is presented in Table 3.3 (labeled "Full Set") and also presented in Figure 3.4. The performance is compared to prior probability for each of the predictions. (Prior probability represents the accuracy of a model that picks the most frequent answer at all times). A comparison shows that all models perform significantly better than the prior probability baseline (for SIA-5 models $G^2(1,3805) \geq 1335$, $p<.001$, for SIA-10 models $G^2(1,3161) \geq 916$, $p<.001$). A comparison of accuracy between models created using the SIA-5 and the SIA-10 data sets revealed no significant differences in accuracy.

Table 3.3 Accuracy (in %) of models compared to baseline by data sets (SIA-5 vs. SIA-10) and prediction class (30secs, 1, 2, 5, and 10 minutes).

	Predict response within	30sec	1min	2min	5min	10min
SIA-5	Full Set	79.8	83.8	87.0	89.4	90.1
	Baseline	54.7	55.9	63.8	72.0	75.4
SIA-10	Full Set	77.5	84.1	86.7	89.6	88.9
	Baseline	54.7	55.1	62.2	70.7	74.2

In order to make sure that the high accuracy achieved by the models is not a result of high accuracy with the more frequent level and poor accuracy with the less frequent level, I present the F-measure produced by the models for each of the levels in Table 3.4. The F-measure is the harmonic mean of each level's *precision* (precision is the percent of predictions of a certain level that were correct) and its *recall* (recall is the percent of a certain level that were predicted as belonging to that level). The high F-measures shown in Table 3.4 for predicting that a user will respond (Yes) as well as for predicting that the user will not respond (No) indicate that the models do an excellent job on both the frequent as well as the less frequent levels.

Table 3.4 F-measures for predictions of response (Yes) and no response (No) by data sets (SIA-5 vs. SIA-10) and prediction class (30secs, 1, 2, 5, and 10 minutes). (F-measure is the harmonic mean of a level's precision and recall).

Predict response within	30sec		1min		2min		5min		10min	
	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
SIA-5	.82	.78	.82	.86	.82	.90	.80	.93	.79	.94
SIA-10	.80	.75	.82	.86	.82	.90	.82	.93	.78	.93

3.8.1 Buddy-Independent Models

In order to understand the role that buddy state and identity play in the predictions, I next examine ten predictive models of responsiveness created after removing all buddy-related features. I will refer to those as *buddy-independent* models.

Buddy-independent models are interesting also as they offer a different solution from a practical standpoint. Models that use the full feature-set (knowing, for example, how much time has passed since the last time a message was exchanged with a specific buddy) may predict, at the same time, different levels of responsiveness to different buddies. In contrast, buddy-independent models are oblivious to information about the source of the message, and will predict, at any point in time, the same level of responsiveness to all buddies, basing the prediction only on information that is "local" to the user.

A comparison of accuracy between the models presented above and the buddy-independent models is presented in Table 3.5. Figure 3.5 shows a graphical comparison for models created with the SIA-10 set.

Surprisingly, while the buddy-independent models performed slightly worse than the models using the full feature set, this difference was not significant. In fact, in some of the models described earlier, the automated feature-selection process selected no buddy-related features even when they were made available. The buddy-independent models performed significantly better than the baseline of prior probability in all cases (for SIA-5 models $G^2(1,3805) \geq 1335$, p<.001, for SIA-10 models $G^2(1,3161) \geq 916$, p<.001). Again, no significant difference in accuracy could be found between SIA-5 models and SIA-10 models.

Table 3.5 Accuracy (in %) of models compared to baseline by data sets (SIA-5 vs. SIA-10), feature sets (Full vs. Buddy-Independent) and prediction class (30secs, 1, 2, 5, and 10 minutes)

	Predict response within	30sec	1min	2min	5min	10min
SIA-5	Full Set	79.8	83.8	87.0	89.4	90.1
	Buddy-independent	79.8	83.7	87.0	89.4	89.3
	Baseline	54.7	55.9	63.8	72.0	75.4
SIA-10	Full Set	77.5	84.1	86.7	89.6	88.9
	Buddy-independent	77.5	84.1	86.6	89.6	88.6
	Baseline	54.7	55.1	62.2	70.7	74.2

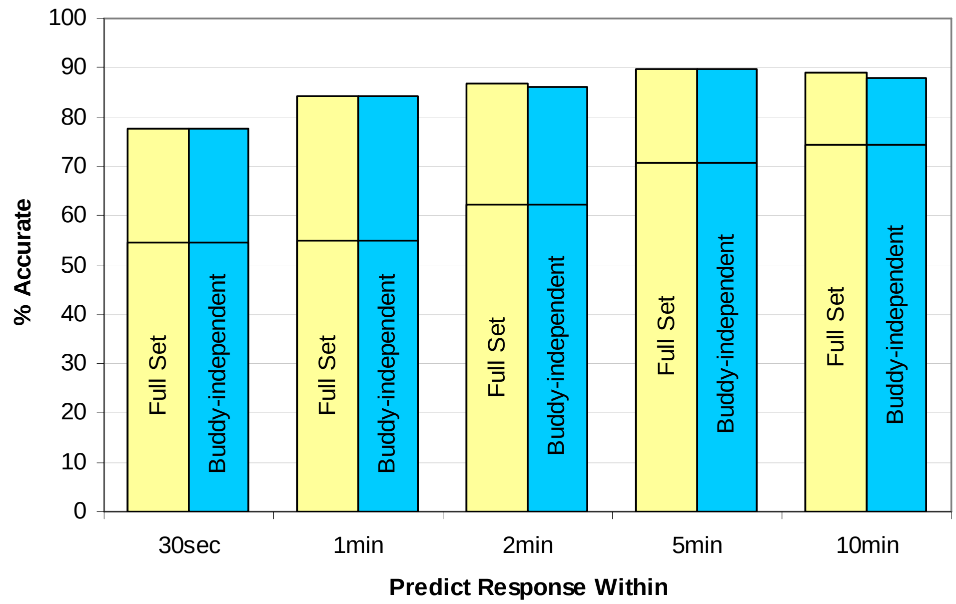


Figure 3.5 Accuracy (in %) of SIA-10 models compared to baseline by feature sets (Full Set vs. Buddy-Independent) and prediction class (30secs, 1, 2, 5, and 10 minutes). Baseline prior probability is shown with the black lines.

3.9 A Closer Look at Selected Features

Following model generation I examined the features that were automatically selected for the 20 models presented above. These features represent those providing the most useful and predictive information to the model. Models built from the full set of features selected on average 12.3 features, while buddy-independent models selected, on average, 10.4 features (this difference is not significant).

3.9.1 Most Selected Features

Since the combined total of distinct features selected by all models was high (57 out of the possible 82), for this discussion I group together features describing similar user activity and application information regardless of the time interval they describe (e.g., group all Keyboard Count features together). I further group features into 3 high-level categories: buddy-related IM information, buddy-independent IM information, and desktop information.

The top 10 selected features for both types of models are:

Full-Data Models	**Buddy-independent Models**
Mouse Distance Traveled (pix)	Mouse Distance Traveled (pix)
Mouse Event Count	Time Since Last Outgoing Message
Time Since Last Outgoing Message	User Input Count
Most Focused Window Type	Most Focused Window Type
User Input Count	Mouse Event Count
Keyboard Count	Duration of Own Status
Time in Most Focused Window	Own Status
Duration of Own Status	Keyboard Count
Time Since Last Incoming Message from Different Buddy	Location (laptop/work/home)
Time Since Last Outgoing Message to Different Buddy	Window Switches Count

Note that the top features selected for both types of models each include six features that are related to desktop activity, (four of which are directly related to user input). This indicates significant predictive influence from the amount of user interaction. Of features related to IM, the time since the last outgoing message, as well as the duration of the current online-status of the participant appear in both lists. It is possible that the duration of status was frequently selected by the models as it could indicate a recent change of state. Finally, we can see that two features describing IM interaction with other buddies were frequently selected for models built from the full set of features for predictors of responsiveness.

3.9.2 Distribution of Feature Types

Next I examined the distribution of feature selection by high level category. On average, full-set models selected 55.3% desktop features, and 44.7% IM features (22.8% buddy-independent IM features, and 22% buddy-related IM features). When moving from these models to buddy-independent models, the distribution of selected features shifts to 62.6% desktop features and 37.4% IM features, suggesting that the void left by the removal of buddy-related IM features was filled, for the most part, by buddy-independent IM features.

3.9.3 Contribution of Desktop Features by Time Window

As described above, desktop features accounted for over 50% of the features selected by the models. The desktop features that were generated looked at different time intervals (e.g., from the last 5 minutes vs. from the last 30 seconds). Figure 3.6 shows the percentage that features with different time intervals were selected for both full-data models and buddy-independent models. It is interesting to observe that desktop-features using longer intervals are selected more frequently, potentially because they provide information that is less susceptible to small changes and noise or because longer trends have more predictive importance.

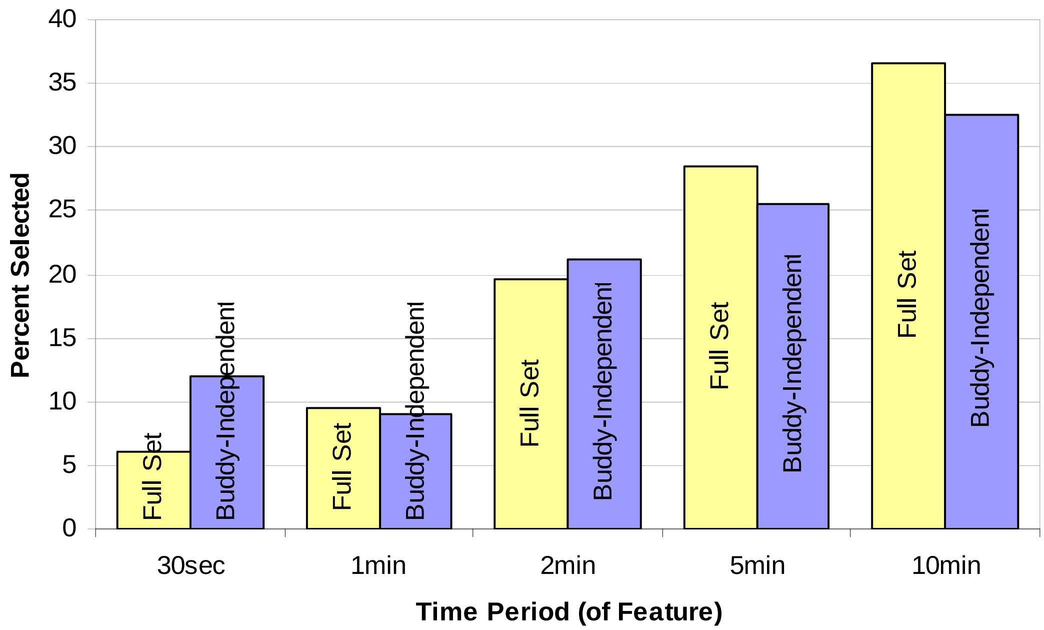


Figure 3.6 Percent of desktop features selected grouped by the time period they were computed on.

3.10 The Use of Old vs. New Training Data: Accuracy Comparison

An interesting practical question regarding the ability to predict responsiveness is that of the possibility to predict responsiveness for one user, with models trained on the data of other users. If such bootstrapping is successful, it would allow systems to provide predictions of responsiveness right as a user begins using them (without first requiring a training period). As time goes by, and with more and more training data collected for the particular user, the system would be able to gradually transition to using models trained primarily on the individual's data. Begole et al. (2003), raise the issue of model inaccuracies as a result of changes in people's routines over time. Thus, it may still be beneficial to use some training data collected from other users, even when a large amount of training data is available for an individual, since, while some states will not have yet been encountered by the individual, it is possible that such states were recorded from other users.

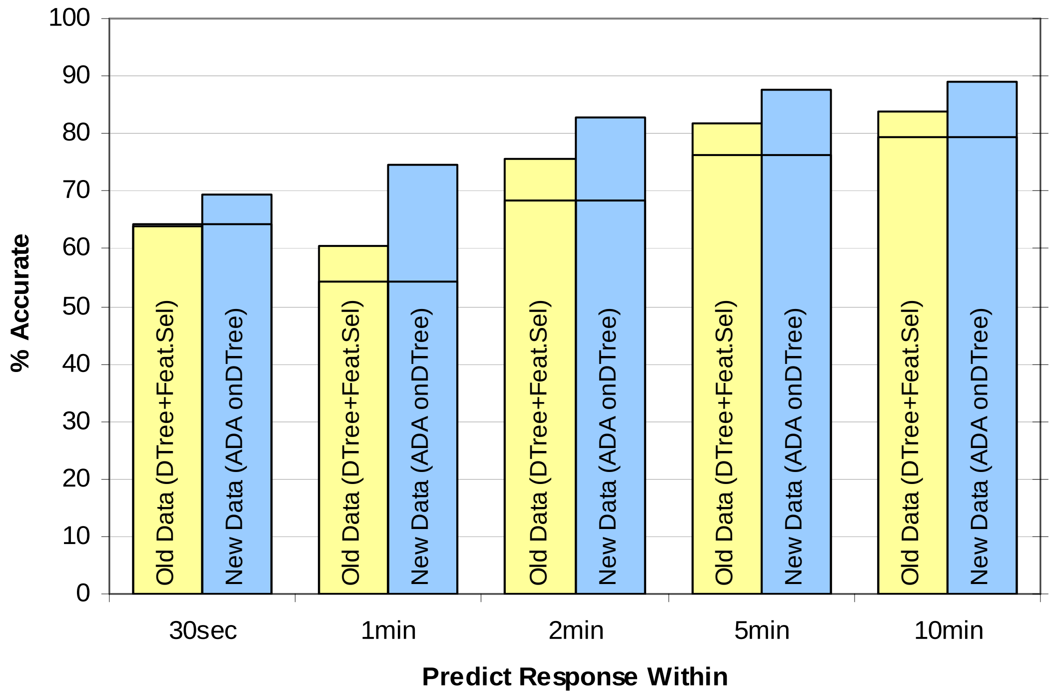


Figure 3.7 Model accuracy when trained on data of Researchers, Intens, and Students groups, compared to models trained on data of the Startup group (all models tested on Startup participants' data).

To compare the accuracy of models created using old data versus new data, I have created the two sets of five models (for the five responsiveness thresholds used throughout this chapter) that are presented in Figure 3.7. The first set of models was created using a wrapper-based feature selection on decision trees and trained on the combined data of the Researchers, Interns, and Students. These models were then tested on the data of the Startup group participants. The second set of models was created using boosting over decision trees using the AdaBoost algorithm. The accuracy of the second set of models was tested using a 10-fold cross validation.

As can clearly be seen in Figure 3.7, the accuracy of models that were trained on old data was lower than that of models trained and tested on the new data. However, both sets of models performed significantly better than the prior probability (with the exception of predictions for the 30-second responsiveness threshold). These results indicate that using old data to train and predict responsiveness of new users can provide significant gains. Especially early on, when a user's behavior has not yet been observed (and the prior probability for them is still unknown), the use of models trained on old data could be most useful.

An interesting observation encountered during the creation of the models described above, was that models trained on old data did better with more general algorithms (decision-trees vs. ADABoosting on decision trees) and parameters (higher minimum of elements per leaf in a decision tree) than with more detailed parameters.

3.11 Discussion

In this chapter I have presented statistical models that are able, with high accuracy, to predict responsiveness of IM users. Specifically, these models are able to predict whether a user is likely to respond to an incoming message within a certain time period. Since the participants in the study showed a high level of responsiveness overall, I was particularly interested in predicting responsiveness to messages that represent a buddy's attempt to start a new session (incoming *Session Initiation Attempts*).

Indeed, predictive models of responsiveness can be applied in a number of useful ways. For example, models can be used to automatically provide different "traditional" online-status indicators to different buddies. Alternatively, models can be used to increase the salience of incoming messages that may deserve immediate attention (such as in Avrahami & Hudson, 2004) if responsiveness is predicted to be low. Models could also be used by a system that will show a list of potentially responsive buddies to users who are looking for help or support, while hiding others. I will now discuss a number of issues regarding the practical use of predictive models of responsiveness.

3.11.1 Implications for Practice

3.11.1.1 Preserving Plausible Deniability

One of the key benefits of IM is users' ability to respond to messages at a time that is convenient to them (or even not respond at all). The insufficient awareness provided by most IM clients is at the source of the problem that we are trying to solve with the models. However, it is the ambiguity inherent in this insufficient awareness that provides users with 'plausible deniability'; that is, it allows them to claim that they did not see a message or even that they were not at their computer. It is thus important to warn against a naïve use of predictions of availability. Providing prediction of responsiveness to buddies "as-is", would substantially reduce plausible deniability and should be avoided. Instead, careful consideration of the application and presentation of predictions is required (for an example of the effect of different awareness displays on timing of interruptions see Dabbish & Kraut, 2004).

3.11.1.2 Making Predictions Visible to the User

In all current IM clients, users can see their own online-status. This allows them to be aware of and control the presence that they expose to others. Similarly, any system providing automatic predictions of responsiveness to others should reflect this information back to the user. One danger, of course, is that users will attempt to learn which factors determine the system's predictions. For example, in a system that uses responsiveness to determine whether to include a user in a set of possible communicators, a user may try to "game" the system in order to always appear as non-responsive. The system, however, can potentially avoid such a situation by making use of predictions from multiple models. A greater number of models, and potentially a greater number of features, could reduce the overall effect of any one feature in the prediction. Finally, allowing users to override the predictions will likely eliminate the need to "game" the system.

3.11.1.3 Multiple Concurrent Levels of Responsiveness

In this chapter I presented a set of models, which I called *Buddy-Independent*, generated using only information about the state of the user without any buddy-related features. My primary reason was to investigate the relative accuracy of buddy-independent models.

However, the use of Buddy-Independent models also has implications for practice. Specifically, a predictive model that takes into account features describing the state and history of a user's interaction with different buddies will, inherently, predict different levels of responsiveness to different buddies. On the other hand models that use only information about the state of the user are guaranteed to provide the same prediction regardless of the identity of the buddy initiating the session. In the design of a system that uses models of responsiveness, the system designer will need to carefully consider whether to provide a unified prediction of responsiveness to all buddies or whether additional benefit may be gained by providing different predictions to different buddies

3.11.2 Content and Topic

One limitation of the models presented in this chapter is that they are unaware of the content of messages sent and received. A large number of messages do not in fact require immediate responses. Avrahami and Hudson (2004) list different levels of responsiveness expected for different types of messages. A model for predicting responsiveness that does not use the content of messages will use other features to explain the lack of a response, potentially leading to inaccurate predictions.

Predictions of responsiveness without using content may also result in misinterpretations of availability. An example of a case where mere responsiveness incorrectly reflects availability is that of responses used for deferral. For example, a user responding quickly with a message saying “can’t talk, in a meeting” would demonstrate high responsiveness but low availability. A model unaware of the content of the message is likely to misinterpret this behavior. In order for such events to be classified correctly they should, more appropriately, be noted in the training data as “no response”. This, however, would be impossible to detect without the content of the messages (and even then, detecting those in an automatic way is not trivial).

3.11.3 Data Size

A second limitation is that the data collected and used for the creation of the predictive models included the logs of only 16 users as they communicated with about 400 buddies. To contrast, Leskovec and Horvitz (2007) present an examination of a data set that contains 300 million conversations between 240 million IM users. Unfortunately their data do not include the richness of detail found in the data presented here, preventing the creation of similar predictive models. Thus, while the data presented in this chapter may be particular to the specific organizations observed and the particular individuals in them, allows us to better understand the possibility for the creation of extremely accurate models of IM interaction.

3.11.4 Other Feature-Subsets Models

In this chapter, I described models created with either all features available, or only buddy-independent features. In future work, I plan to compare the accuracy of models created using different feature subsets. In particular I am interested in the predictive accuracy of models generated using only IM features -- features that describe the user’s communication state. If these models, created without the use of desktop features, would have accuracy that is comparable to accuracy of the models presented earlier, then these new models would present a much simpler and elegant solution for the construction of predictive models of responsiveness.

3.11.5 Beyond Desktop Events

Previous work (for example, Horvitz et al., 2002; Begole et al., 2004; Fogarty et al., 2004a) described the creation of statistical models that used input from a person's calendar as well as sensors external to the workstation. Those included a door sensor, sensing whether the door was open or closed, a phone sensor, sensing whether the phone was on or off hook, simple motion detectors, and speech sensors, implemented with microphones installed in the person's office, or the microphone built into participants' laptops (Fogarty, Au, & Hudson, 2006, attached microphones to water pipes for doing simple activity recognition). When designing the data collection for the work presented in this chapter I decided not to use sensors external to the desktop. While I believe that it is reasonable to expect events and activities external to computer usage to be reflected in that usage (for example, a user attending to a visitor is likely to generate fewer computer events), I suspect that improvement to the models could potentially be generated from features that use such sensor data. Fogarty and Hudson (2007) presented a toolkit representing an effort for reducing the difficulty associated with the collecting of data and generation of predictive models (their tool, however, does not currently support responsiveness as a valid label for learning).

CHAPTER FOUR

Forecasts of Responsiveness

In the previous chapter I have discussed the benefit that would come from predicting responsiveness to a message *before* that message is sent. Such a prediction can allow a user to decide whether or not to initiate communication with a buddy. In this chapter, I examine the need for predicting responsiveness from a different angle – the likelihood of responsiveness to a message *after* that message was sent. Consider the case where a user has already sent a message and is now waiting for a response (this message could, but does not have to be a session initiation attempt). This user may wish to know the likelihood that a response will (or will not) arrive within some period given that they have already been waiting for some time (e.g., the likelihood that they will receive a response within 5 minutes given that they have already waited 5 minutes for a response). This may allow them to decide whether to wait longer or tend to other matters. Alternatively, the user may wish to know how long they will have to wait in order to receive a response with some likelihood (for example, it might be useful for the user to know, having already waited for two minutes, that they will need to wait another 25 minutes in order to have a 50-50 chance of getting a response.)

In related work, as part of Priorities research on presence forecasting, Horvitz et al. (2002) explored the influence of time away on cumulative distributions for return to a desktop computer (that is, the likelihood that a user will return to their main desktop system, given that they have been away for different periods of time). Similar to the investigation of rhythms of presence by Begole et al. (2003) and to the investigation reported here, Horvitz et al. also examined the difference in presence-forecasting during different times of the day. In an investigation of task switching and resumption in the face of interruptions Iqbal and Horvitz (2007b; , 2007a) presented the influence of time away from a task on the cumulative probability of that suspended task being resumed. They further examined the effect of the user's interaction with the interruption (in this case an instant message, email, or conversation) on this cumulative distribution.

4.1 Estimating the likelihood of a response

In instant messaging, as well as in other asynchronous communication mediums, a user who has sent a message and is waiting for a response may wish to find out the likelihood of a response to their message given the time that has passed since they sent their message (and they may wish to query this likelihood again after an additional wait period with no response). This likelihood can be defined as the conditional probability

$$p(\ t_r < t_{w2} \mid t_{w1}\ )$$

where t_r is the time to receive a response, t_{w2} is the additional wait time, and t_{w1} is the time period already waited.

In order to provide this estimate of the likelihood of a response to an incoming IM, I have examined the timing and probability of responses to messages in my data using both incoming messages (sent to my participants by their buddies) and outgoing messages sent by my participants. This allows examining the likelihood of responsiveness by my participants as well as the 412 buddies present in the data (although responsiveness by the buddies is only available for their communication with the participants). I have specifically examined the likelihood of a response following a wait period by excluding messages that were followed by messages from the same sender. For example, if a buddy sent a message and then sent a second message some time after (without a response in between), only the second message was included.

Figure 4.1 shows a set of smoothed curves representing the probability of receiving a response within different time periods, given that a response has not arrived for different wait periods. Using the probability formula described above, Figure 4.1 shows the corresponding wait time t_{w2} for a set of desired probabilities (.05, .1, .25, .5, .75, .90, and .95).

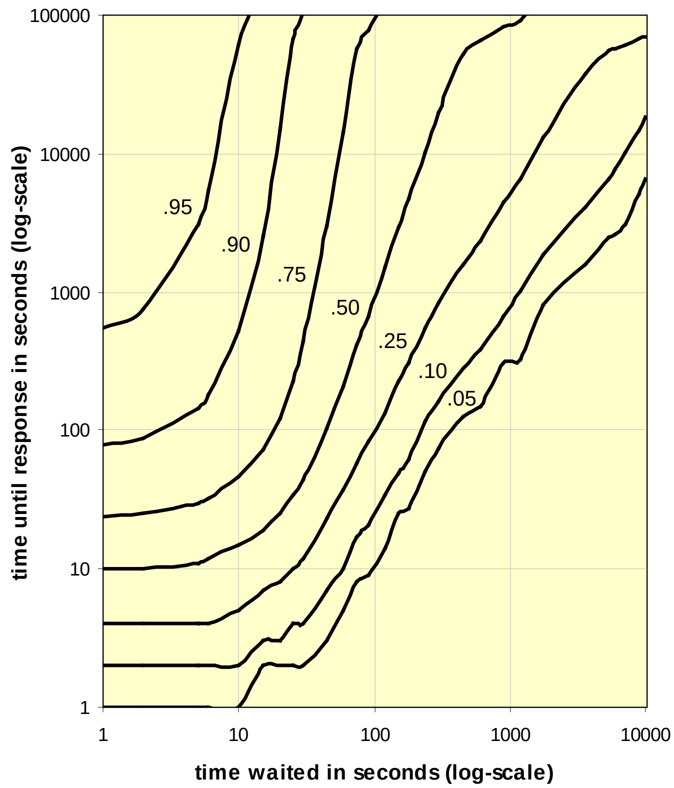


Figure 4.1 Likelihood of time until response at different probabilities given time already waited.

Figure 4.2 displays curves representing the likelihood, given that a response has not arrived for a period of time, that the response will arrive if an additional period of equal length is waited. Using the formula described above, what is the probability p such that $p(t_r < t_{w2}{=}t_{w1} \mid t_{w1})$. The darker curve represents the overall likelihood of a response, considering the amount of time since the message was sent. The dashed curves represent likelihood of a response during different parts of the day (Morning, Lunch, Evening, and Night*). For

* Morning: 6:00-11:30, Lunch: 11:30-14:30, Evening: 14:30-18:00, Night: 18:00-6:00

example, the graph shows that in 42% of all cases examined, if a response has not arrived for 30 seconds it will arrive in the following 30 seconds. Interestingly, as can be seen in Figure 4.2, the likelihood of receiving a response following a delay is higher at night than during the day.

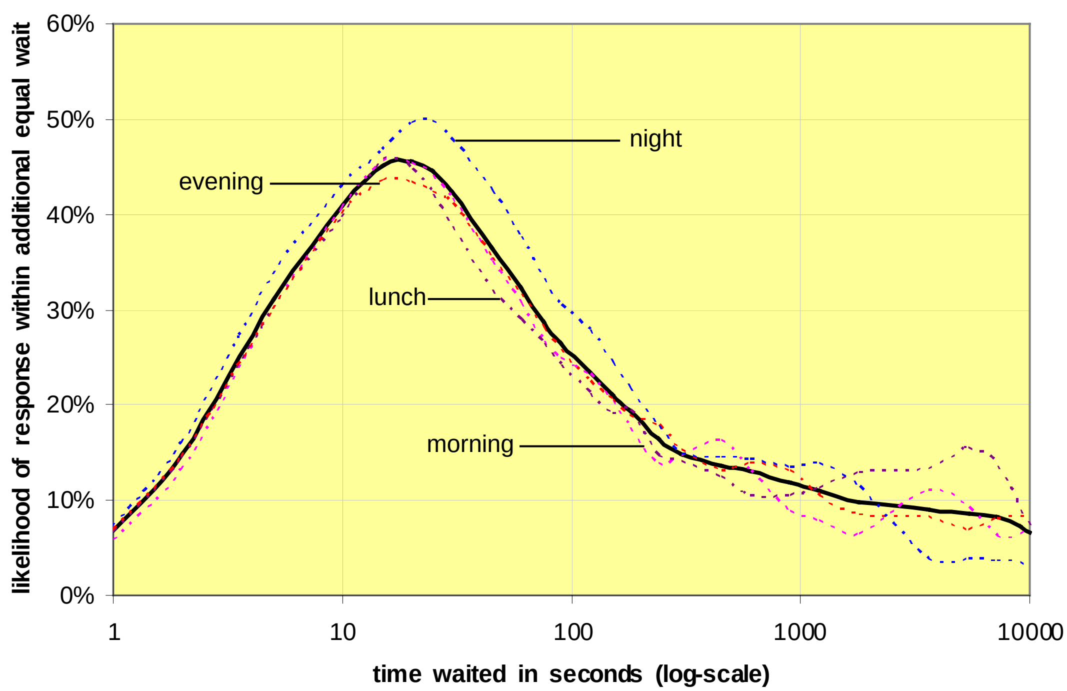


Figure 4.2 Probability of a user replying within the same time period already waited (e.g., respond within 15 minutes given that 15 minutes have passed). Shows probability for all data (bold) and for the data segmented by part of day.

Figure 4.3 shows the likelihood of a response arriving within some time period. Each curve represents a response arriving within a specific time period. For example, the analysis shows that after waiting for a response for 5 minutes, the likelihood that a response will arrive in the next 10 minutes is 23% while the likelihood of a response in the next half hour is 35%.

From the opposite perspective, the graph presented in Figure 4.3 could be considered to present the likelihood that a session has ended, taking the probability that a response will not arrive within some time period to be 1 minus the probability of a response within that period. For example, given that, having already waited for five minutes, the probability that no response will arrive in the next 10 minutes is 77%, a user may consider that the IM session has likely ended.

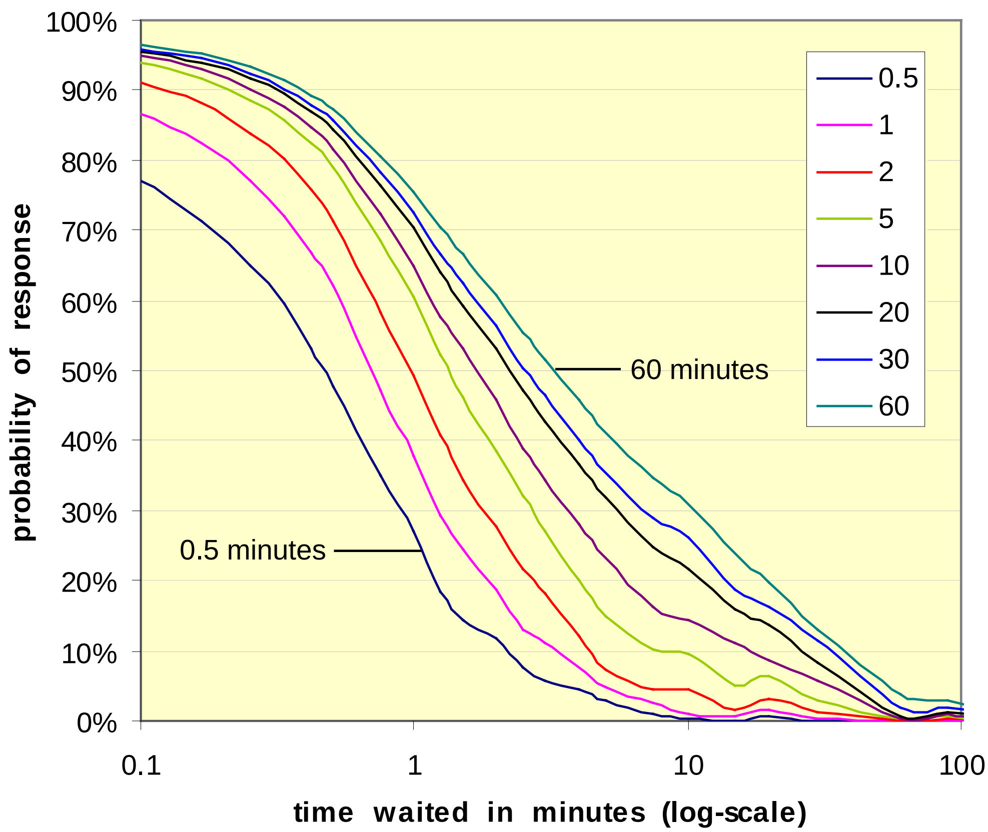


Figure 4.3 Probability of a user replying within 0.5, 1, 2, 5, 10, 20, 30, and 60 minutes, conditioned on the time period already waited.

Finally, forecasts of responsiveness may also be performed at the level of the individual buddy. Figure 4.4 illustrates the probability of receiving a response within two minutes, from two different individuals (for demonstration, I have chosen the two participants who were slowest and fastest to respond). Compare, for example, these individuals' likely responsiveness after a wait of 1 minute. As can be seen in their likelihood distribution, the fast responder is 50% likely to respond in the following two minutes, while there is only 19% likelihood of receiving a response from the slow responder. Such forecasting for different individuals, if sufficient data are available for them, may prove most beneficial.

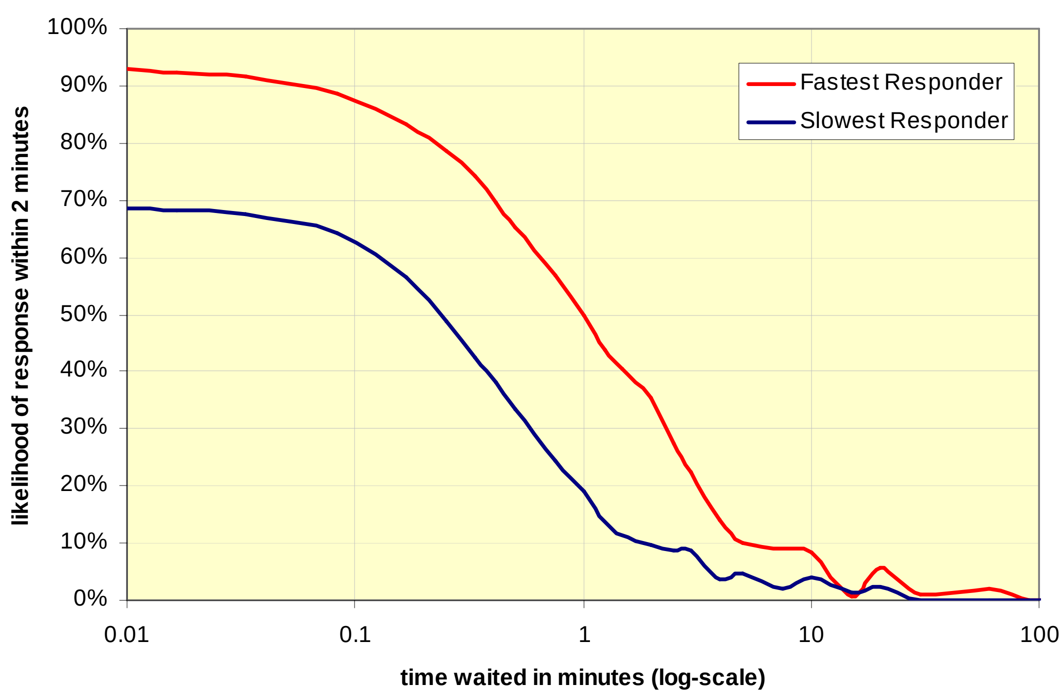


Figure 4.4 Probability of a reply within 2 minutes from two different buddies (slowest and fastest respondents in my data), conditioned on the time period already waited.

4.2 Conclusion

In this chapter I presented an investigation of the likelihood of responses in situations where the message has already been sent and the user has been waiting for a response for some period of time. The distribution of interactivity and responsiveness in email and online forums have been previously studied (Kalman, Ravid, Raban, & Rafaeli, 2006a, 2006b). Similar to the data presented in this paper, asynchronous communication has been shown to typically follow a Power Law distribution. Barabási (2005) proposed a theoretical priority-based task-selection model that allows the explanation of these observed distributions. The investigation of response-likelihood presented here, however, may provide benefit beyond merely adding to the body of research on the probability distribution of asynchronous communication, rather providing multiple different and potentially applicable views into the underlying distribution of IM responsiveness, similar to that provided in the Priorities and Coordinate systems (Horvitz et al., 2002).

CHAPTER FIVE

Understanding IM Responsiveness

5.1 Introduction

When faced with incoming communication, one must quickly weigh a multitude of factors in order to decide whether or not to engage in the communication. Similarly, when deciding whether or not to initiate communication, one will often try to estimate the other's availability for the communication to assist in the decision.

Dabbish (2007) presented a model outlining many of the factors that affect a receiver's choice to engage in communication, including the cost of postponing one's primary task, the perceived benefit of the communication (to one's self and to the initiator independently), and the ongoing relationship with the initiator. Through a series of laboratory studies she was further able to assign numerical weights describing the interplay between these factors.

In my work I have argued for the importance of the concept of *responsiveness* as one of the few observable behaviors through which we can sense and even predict another person's *availability* to communication, referring to responsiveness as *demonstrated availability*.

Responsiveness to communication may indicate not only how important an incoming communication is perceived to be, but also the user's level of engagement in whatever task they were engaged in prior to the communication. This argument is supported by the work by Tyler and Tang (2003). In their interviews they found that email users would sometimes change their responsiveness when replying to emails intentionally and consciously, for the purpose of conveying to the sender their availability (as well as projecting their perception of the level of importance of the communication).

In Chapter 3, I have presented a set of highly accurate predictive models of responsiveness to incoming instant messages. These models were created from data that were collected in an unobtrusive fashion and without requiring user labeling. I have then described a diverse set of applications that may enhance communication through the use of such predictive models of responsiveness. However, it is clear that a better and deeper understanding of responsiveness is still needed; how does the user's ongoing activity (or activities) affect their responsiveness to incoming (and potentially interrupting) communication? Will responsiveness be different to different people? Will they respond at different speeds during different parts of the day? How will the content of the communication affect the user's responsiveness to it? How will responsiveness, when the communication is already ongoing, differ from responsiveness to new communication?

In this chapter, I present a careful look at responsiveness to IM, through an in-depth quantitative analysis, in an attempt to answer the following research question:

How do context, communication and content, affect a user's responsiveness?

I present, for example, findings that show that work-fragmentation, or a user's frequent transition between applications, significantly correlates with faster responsiveness. I show also that the salience of an incoming message has significant effect on responsiveness – even greater than indicators that an incoming message is part of ongoing communication.

It is my hope that through the results of this analysis I am able to shed light on the concept of responsiveness to communication and its connection to availability, and able to hint at ways in which responsiveness could be influenced.

5.1.1 Responsiveness and Context

Similar to an incoming phone call, an incoming instant message finds the user in some particular context that may affect their responsiveness to the communication. Furthermore, context may affect responsiveness such that it changes from message to message within the same conversation. (In phone calls, responsiveness is mostly interesting in the time to initially accept the call.) Multitasking when engaged in a phone call or face-to-face conversation can be difficult or inappropriate since high responsiveness is usually expected and delays are quickly noticed and meaning is often attributed to these delays (see, for example, Schegloff & Sacks, 1973; McLaughlin & Cody, 1982). Unlike with phone calls, however, the semi-synchronous nature of IM allows users to easily multitask while engaged in communication by using breaks between conversation turns to resume other tasks or attend to other IM communication. (This is not to say that delays in responses in IM go unnoticed.)

In the work described here, I examine how responsiveness to incoming messages is affected by the user's context, looking specifically at the user's other ongoing computer activities, their other recent and ongoing IM activity, and global context including the day of the week and time of day.

5.1.1.1 Responsiveness and Desktop Context

As mentioned above, an incoming message may find a user engaged in many different activities (both on and off the computer). Even when looking at context strictly as represented by the user's activities on the computer, the user may be in greatly varied contexts. For example, the incoming message may find the user engaged in a complex programming or design task that requires their attention, or may find them using the computer for messaging and other communication, or simply for listening to music. Work by Iqbal and Horvitz demonstrated that users exhibit an increase in behaviors associated with task suspension before switching to an incoming email or IM (Iqbal & Horvitz, 2007b), or voice communication (Iqbal & Horvitz, 2007a). These behaviors include an increase in document saves, paragraph completion, etc. One might expect the type and complexity of a task, as well as the amount of work required to leave that task in a "stable" state before switching away to an incoming message may affect the responsiveness to that message. As described in Section 3.7, the measures that I used for describing a user's desktop context include the amount of keyboard and mouse activity, the amount of window switching, and the type of the application most used prior to the arrival of the message.

5.1.1.2 Responsiveness and IM Context

Due to the semi-synchronous nature of IM, users will often find themselves engaged in more than one IM conversation in parallel. However, high levels of responsiveness to simultaneous communication may be difficult to sustain (and may result in the user feeling overwhelmed). I conjecture that such simultaneous ongoing communication will significantly reduce users' responsiveness to incoming communication.

While the presence of other ongoing communication may reduce one's responsiveness, the *recency* of communication with others may be an indication of the user's receptiveness to communication. This may thus suggest that recent IM communication with others could be associated with faster responsiveness.

5.1.2 Responsiveness and the Communication Partner

A number of elements associated with the sender of the incoming message may affect a user's responsiveness to that message. For example, the specific identity of the sender or the type of relationship the user has with this buddy may affect responsiveness. Furthermore, the identity of the buddy may have a different effect when deciding whether to engage in conversation compared to when a conversation is already ongoing. The time that has passed since the previous communication with this particular buddy may also affect responsiveness. On one hand, recent communication may suggest that the user will be fast to respond to further communication. On the other hand, users may be interested and curious about incoming communication from buddies, with whom they have not communicated for a long time. Examining the effect of measures of the buddy identity and communication with the buddy will hopefully shed light on the surprisingly high accuracy of the buddy-independent models presented in Chapter 3.

5.1.3 Responsiveness and Content

Finally, the content of the message and the content of the conversation to which it belongs are sure to have an effect on responsiveness. In related work, Burke, Joyce, Kim, Anand, Kraut (2007) showed that different linguistic features, extracted from Usenet messages, significantly correlated with the likelihood of these messages receiving a response. Dabbish, Kraut, Fussell, and Kiesler (2005) found that different elements of the content of email messages significantly correlated with the importance that users attributed to incoming messages. This, in turn, significantly affected the likelihood that they would respond to the incoming email. As discussed by Avrahami and Hudson (2004), different messages are associated with different expectations of responsiveness; while some may require an immediate response, other may allow a leisurely response, or require no response at all.

While a detailed examination of the content of messages and their relation to responsiveness is left for future work, a number of potentially relevant attributes have been extracted and examined in this current work. The first and most basic of these measures is the length of the incoming message. One may expect the length of an incoming message to have a significant effect on responsiveness. This is not to suggest, however, that it is the length of the message *per se* that causes this effect, rather that other factors that are manifested in message length (such as the complexity of the content, or the courtesy of the communication) have an effect on communication. Avrahami and Hudson (2006a) showed that relationship between a user and a buddy has a significant effect on the length of the messages exchanged (with significantly longer messages exchanged between buddies in a work relationship compared to buddies in a social or a combination of work and social relationship). Isaacs et al. (2002) showed the effect of user's frequency of use of IM on the length of messages. The other content-related measures that were coded are whether the message contains a question, whether the message contains a URL, and whether or not it contains an emoticon. (Emoticons are combinations of characters, such as the famous :-) smiley face, that are often used in chat to express emotion.)

5.2 Outline

Since the analysis described in this chapter involved a few subtle steps, I have elected to describe the results of a small number of sub-analyses before describing the analysis method in detail. This, I hope, will make the reading of this chapter easier, more understandable, and hopefully more interesting.

In the remainder of this chapter, I describe the full list of measures that were investigated followed by a description of a number of steps taken to prepare the data for analysis. I then describe basic and important findings that influenced the final analysis that was performed. The first shows a significant relationship between responsiveness and a user's 'online-status'. The second shows a significant effect of the state of a message window prior to the arrival of the message on responsiveness. These two findings are followed by a detailed description of the analysis method and the findings of the analysis. A discussion of the results concludes this chapter.

5.3 Measures

The set of measures examined in this chapter were computed from participants' logs. The measures are grouped into 3 high level categories: Context, Communication, and Control.

Context Measures

These measures represent the context into which an incoming message arrives and include global context (such as the time of day), the participant's other ongoing desktop activities and other IM communication:

Global Context

- Day of the Week (Monday through Sunday)
- Part of Day (Morning, Lunch, Evening, Night)*

IM Context

- Online Status (Online, Idle, Be Right Back, Away)
- Length of time in current online status (log-transformed)
- Whether there are other IM windows open (Single Window vs. Multiple Windows)
- Time since the last message sent to a different buddy (log-transformed)

Desktop Context

- Number of Window-Title Switches including both switching between different applications as well as changing between documents or web-pages in the same application (Principal Component Analysis (PCA) on log-transformation – see details in Section 5.4.5)

* Morning: 6:00-11:30, Lunch: 11:30-14:30, Evening: 14:30-18:00, Night: 18:00-6:00

- The amount of Keyboard activity prior to the arrival of the message (PCA on log-transformation – see details in Section 5.4.5)
- The distance (in pixels) traveled by the mouse pointer (PCA on log-transformation – see details in Section 5.4.5)
- The type of the application that was most in focus in the two minutes prior to the arrival of the message* (Browser, email, word processing, IM client, presentation, etc.)

Communication Measures

These measures describe a number of basic elements of the incoming message as well as elements relevant to the sender of the message.

- The buddy (Buddy ID)
- The relationship with this buddy as indicated by the user (See Table 6.2)
- Time since the last message the user sent to this buddy (log-transformed)
- Time since the last incoming message the user received from this buddy (log-transformed)
- The state of the message window (Existing-Focused, Existing-Not Focused, New-Popup, New-Minimized, New-Hidden)
    - Window (New vs. Existing) (see details in Section 5.5.2)
    - Focus (In Focus vs. Out of Focus) (see details in Section 5.5.2)

* Similar measures for different time periods were computed. However, in order to avoid singularity, only one of these measures could be included in the analysis.

Content

- The length of the message, in characters (log-transformed)
- Does the message contain a question (0 vs. 1)*
- Does the message contain a URL (0 vs. 1)
- Does the message contain an emoticon (0 vs. 1)

Control Measures

These measures represent elements that are constant to each of the participants during their participation period. They are:

- Group (Students, Researchers, Interns, Startup)
- Participant ID
- Gender (Female vs. Male)
- Age

5.4 The Data

Before beginning the analysis, a number of steps were necessary to ensure that the analysis is done correctly and provides the most informative results.

* For a description of the rules used to identify questions see Ch. 0 and Avrahami and Hudson (2004).

The data analyzed included the data from the Researchers, Interns, Students, and the Startup group. Since I had complete knowledge of desktop and IM state of the participants but not of their buddies, the analyses described next examined only responsiveness to incoming messages (rather than also examining buddies' responsiveness to participants' outgoing messages).

5.4.1 Accounting for Differences in Duration of Participation

Since the data collected represent naturally occurring IM interaction, different message volumes were recorded from different participants. Furthermore, some participants elected to continue their participation beyond the required four weeks, again, resulting in differences in the amount of data logged from different participants. One could consider two approaches to help avoid a situation in which the data of a small number of participants overwhelm the results of the analysis. The first approach would be to try and examine similar numbers of messages from the different participants. In the second possible approach, similar data collection periods from the different participants would be analyzed. Following the latter approach, I have decided to use only those data recorded from each participant during the first 45 days of their participation. As such, the average participation period was 36 days (Min=17, Max=45, SD=8.96) with a total of 73571 messages (37547 incoming and 36024 outgoing messages).

5.4.2 Normalizing Measures

Unlike the predictive models described earlier, which provided predictions on five binary classes of responsiveness (within 30-seconds, 1, 2, 5, 10 minutes), in this chapter responsiveness is analyzed as a continuous measure. However, the time until a response, as well as a number of the explanatory measures used (for example, the time since receiving a message from a different buddy), exhibit a non-normal distribution with a peak and a long tail. For example, as mentioned earlier, 92% of the incoming messages were responded to within 17 seconds, and 50% of the messages were responded to within 5 minutes (similar responsiveness distribution was reported by Kalman et al., 2006a).

To address this issue I have used a log-transformation on these measures. In order to keep the results interpretable, a log base 10 was used.

5.4.3 Handling Non Response

A number of incoming messages in the logs (240 messages, to be exact) were never responded to by a participant. That is, the participant did not send an outgoing message to the same buddy before completing their participation in the study. For the purpose of analysis, I assigned to these messages a responsiveness value that is equal to the number of days of participation remaining for the participant (since it is possible that a response was sent right after the end of the participation).

5.4.4 Accounting for Dependence between Messages

When analyzing the logs, or other data of such nature, we must keep in mind that messages arriving in close time proximity are not independent of one another. That is, two different messages arriving one after the other (even if they were sent by different buddies) are likely to find the user in a very similar state and to result in similar responsiveness to these two messages (or "what happened just before will likely happen now").

Indeed, in my data, the responsiveness to a message was highly correlated with responsiveness to the previous message. The correlation between two consecutive messages from the same buddy was $r=.454$ and the correlation between two consecutive messages received from any buddy was just slightly lower $r=.453$. An Autoregressive analysis (AR), treating the incoming messages as time-series events also revealed a significant correlation between consecutive messages. Thus, in order to account for this lack of independence between the consecutive data points, I included in the analyses described next the user's responsiveness to the previous incoming message from the same buddy (often referred to as the "lag-1") as a control measure.

5.4.5 Reducing Measure Covariance with Principal Component Analysis

The measures of computer activity were computed for a set of time-periods prior to the arrival (or sending) of a message. Specifically, computed measures that describe the number of window-title switches, mouse movement, keyboard activity, and the most used application for each of five time-periods: 30, 60, 120, 300, and 600 seconds preceding the arrival (or dispatch) of the message. As expected, however, the correlation between the measures computed for different windows is very high. For example, the correlation between keyboard activity in the 30 seconds prior to the arrival of a message (log-transformed) and keyboard activity in the 60 seconds prior to the arrival of a message (log-transformed) is r=.82.

```
WinTitleSwitchesPCA =  0.04252 * WinTitleSwitches30secs  +
                       0.08867 * WinTitleSwitches60secs  +
                       0.18429 * WinTitleSwitches120secs +
                       0.47321 * WinTitleSwitches300secs +
                       0.85583 * WinTitleSwitches600secs +
                       (-21.4658)
```

Figure 5.1 An aggregate measure of window-title switches produced using a Principal Component Analysis on covariance (first component accounting for 90% of covariance).

In order to prevent the covariance of these individual measures from aversely affecting the analysis, I created three new measures "summarizing" title switches (WinTitleSwitchesPCA), mouse movement (log transformed) (MouseDistancePCA), and keyboard activity (log transformed) (KBCountPCA). This was done by conducting Principal Component Analysis (PCA) three times, keeping the first component from each. For example, the five measures describing window-title switches (for 30, 60, 120, 300, and 600 seconds) were combined through linear combination into a single measure WinTitleSwitchesPCA (see Figure 5.1 for the specifics of the linear combination).

5.5 Initial Findings

Before describing the full analysis of responsiveness, I start by presenting two basic and important findings. The first finding shows a significant relationship between responsiveness and a user's 'online-status'. While seemingly intuitive, confirming this finding allowed me to focus my remaining investigation on cases for which the participant did not explicitly indicate unavailability or was inactive. The second finding shows a significant effect on responsiveness of the state of a message window prior to the arrival of the message. This finding led to the examination of responsiveness separately under different message window states.

5.5.1 Online Status and Responsiveness

Indicators of presence and explicit indications of availability are one of the unique and most important features of Instant Messaging. Through the *online status* of a buddy, users can tell, before initiating communication, whether a buddy is online and present at their computer, whether they have been inactive for some time, or even whether they have indicated themselves to be occupied or busy. It is important, however, to distinguish between indicators of presence and indicators of availability. That is, a user who is present and working on a computer may not be available for communication. As previously noted, this important distinction between presence and availability is too often blurred and ignored (see, for example, Begole et al., 2004; Fogarty et al., 2004b).

The field data that I collected included messages arriving when participants were in different states of presence and availability. While my particular interest is in responsiveness in the absence of indicators of inactivity or unavailability, it was necessary to confirm, first of all, that responsiveness 'behaves' as expected under different online statuses. Thus, as a first step, I examined whether the online status of a participant has an effect on their responsiveness.

The four online statuses used by the participants included *Online*, *Idle*, *Away*, and *Be-Right-Back (brb)*. An *Online* status indicates that the user is connected, present, and has been using the computer recently. *Idle* is set automatically by Trillian after 5 minutes of mouse and keyboard inactivity. The *Away* status is either explicitly set by the user or is set automatically by Trillian following 20 minutes of mouse and keyboard inactivity. Finally, *Be-Right-Back* is set explicitly by the user. It is quite reasonable to expect the different online statuses, representing both explicit indications of unavailability and automatic indications of presence (or lack thereof) to be associated with different levels of responsiveness (e.g., a user's inactivity on the computer would naturally result in slow responsiveness to incoming messages). My hypothesis is as follows:

H1: Users will have significantly different responsiveness to incoming messages when in different online statuses with slower responsiveness when in statuses of explicit unavailability or statuses indicating inactivity.

To examine the effect of online-status on responsiveness and test this hypothesis, I conducted a simple mixed model analysis in which responsiveness (log-transformed) was the dependent measure and Online-Status was the main independent measure of interest. The lag-1 of responsiveness, the control measures (Group, Participant ID, Gender, Age) and global context measures (Day of the Week, and Part of Day) were included as fixed effects. ParticipantID was nested in Participation Group and modeled as a random effect.

The analysis showed that a user's status indeed has a significant effect on responsiveness ($F[3, 37439]= 43.1$; $p<.001$; see Figure 5.2). A Tukey HSD pair-wise comparison showed each of the four statuses to be significantly different from one another. As expected, explicit indications of unavailability significantly correlate with users' slower responsiveness. Extended periods of inactivity (or periods when the user is not present) represented by the *Idle* status, also correlate with slow responsiveness. Responsiveness was fastest when the user was in an *Online* status (M= 47.3 seconds) followed by responsiveness when in an *Away* status (M= 60.3 seconds). Next is responsiveness when the user is in a *brb* status (M=88.8 seconds) and finally, with a significantly longer delay before responding, was responsiveness when the user was *Idle* (M=308.4 seconds). These findings confirm hypothesis 1.

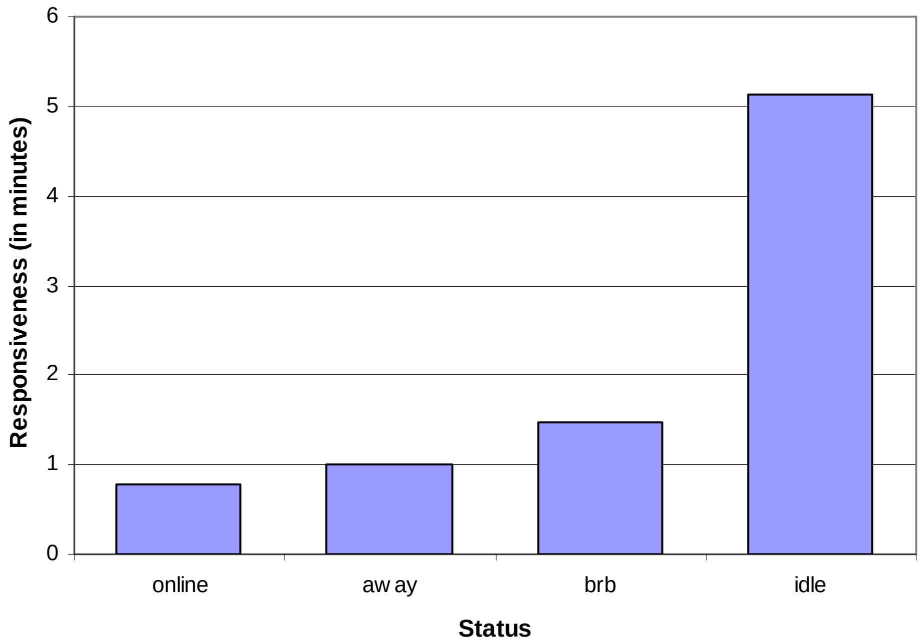


Figure 5.2 The effect of the user's status on responsiveness. The idle status is set automatically after 5 minutes of inactivity. The brb status is set manually by the user. The away status can either be set manually or is set automatically after 20 minutes of inactivity. The analysis found each of the four levels to be significantly different from the others.

It is interesting to note that responsiveness in the *Idle* status is slower than in the *Away* status even though the *Away* status, when automatically set by Trillian, indicates 20 minutes or more of inactivity, compared to 5 to 20 minutes of inactivity represented by the *Idle* status. It is possible that this is due to the fact that the *Away* status can also be manually set by users who do not wish to be disturbed.

Following this initial analysis we may now wish to revise our original research question to ask

> *How do context, communication and content, affect a user's responsiveness when they are present and have not indicated themselves to be unavailable?*

In order to answer this question, in all the analyses described in the remainder of this chapter, I examined only those incoming messages arriving when a participant was in an *Online* status (32399 messages, or 86% of all the incoming messages).

5.5.2 The State of the Window (and User Preference)

Next, I present a second basic yet important finding; that of the effect of the state of the message window, prior to the arrival of the message, on responsiveness.

5.5.2.1 Messages arriving in an existing window

It is reasonable to assume that responsiveness to an incoming message will depend on whether the message-window was already open – a message window that is already open will most likely indicate that some communication with the buddy has previously started. (Note that it is possible that this communication includes only outgoing or only incoming messages.) However, an open window can be either "in focus" as the currently active application – I will refer to this state of the message window as *Existing Focused* – or it can be "out of focus", that is, not the active application, in which case its taskbar icon will flash – I will refer to this state as *Existing Not Focused*.

The *Existing Focused* window state may represent the strongest indication that the user is already engaged in communication with the buddy.

5.5.2.2 Messages arriving in a new window

If, however, the message window was not already open, then its method of appearance depends on the user's preference. In this study, users may elect to have message windows appear in one of three methods:

The *New Popup* Window State: In the first method, upon the arrival of an incoming message, if a message window does not already exist, it is automatically created and displayed on the desktop in the foreground, on top of all other applications. I shall refer to this state as *New Popup*. The *New Popup* method of displaying incoming message is the default presentation in Trillian. It was used by nine of the participants.

The *New Minimized* Window State: In the second method, preferred by eight of the participants, the message window is automatically created but appears minimized on the user's taskbar (see Figure 5.3). I shall refer to this window state as *New Minimized*. In this presentation method, the user is notified of the incoming message through the flashing of the window's taskbar icon. In this *New Minimized* method, the user must then click on the taskbar icon in order to bring the message window to the foreground. (Two of the eight participants who elected this style of message delivery had switched to it after briefly using the default *New Popup* method.)

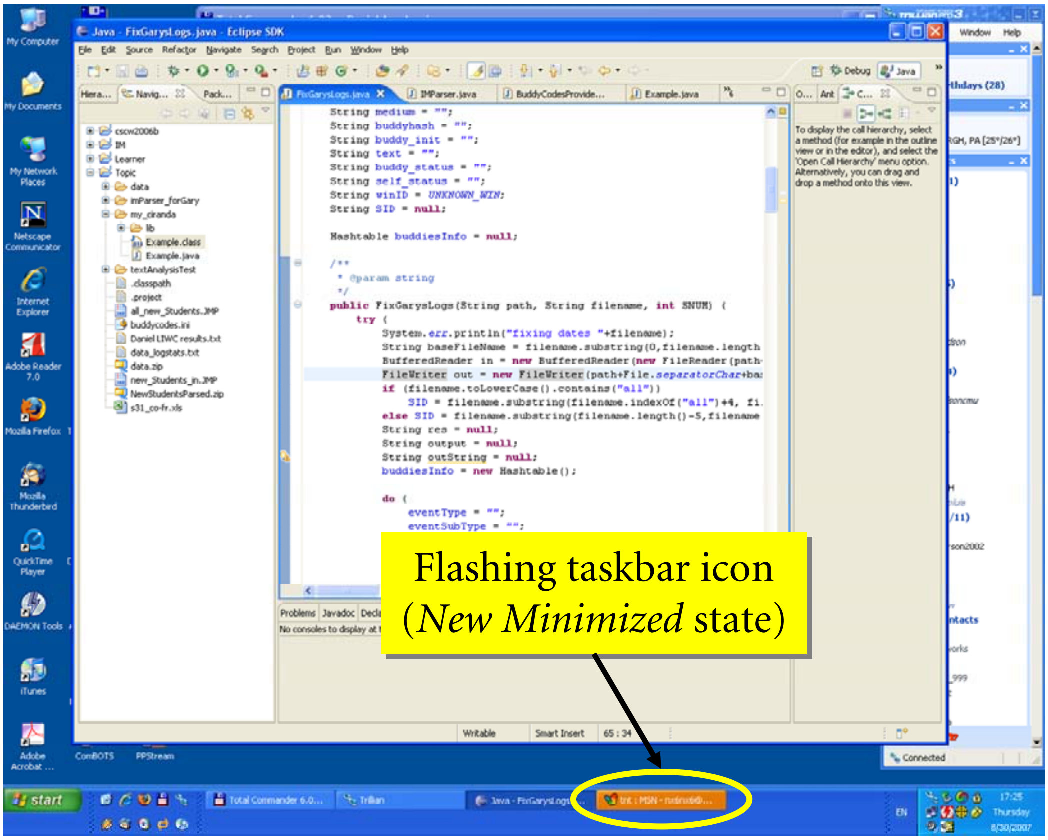


Figure 5.3 Notification of a new message in the *New Minimized* window state through flashing of the window's taskbar icon.

The *New Hidden* Window State: In the third and final method, which I refer to as *New Hidden*, the user is notified of the incoming message through a small (16x16 pixels) blinking icon in the corner of their screen (see Figure 5.4) and a similar icon in their buddy-list (note that the buddy list is very often obscured by other applications). The user must then click on either one of these small icons in order to make the message window appear on the desktop. This message delivery option was preferred by two of the participants while a third participant used this method briefly before settling on the *New Minimized* delivery method.

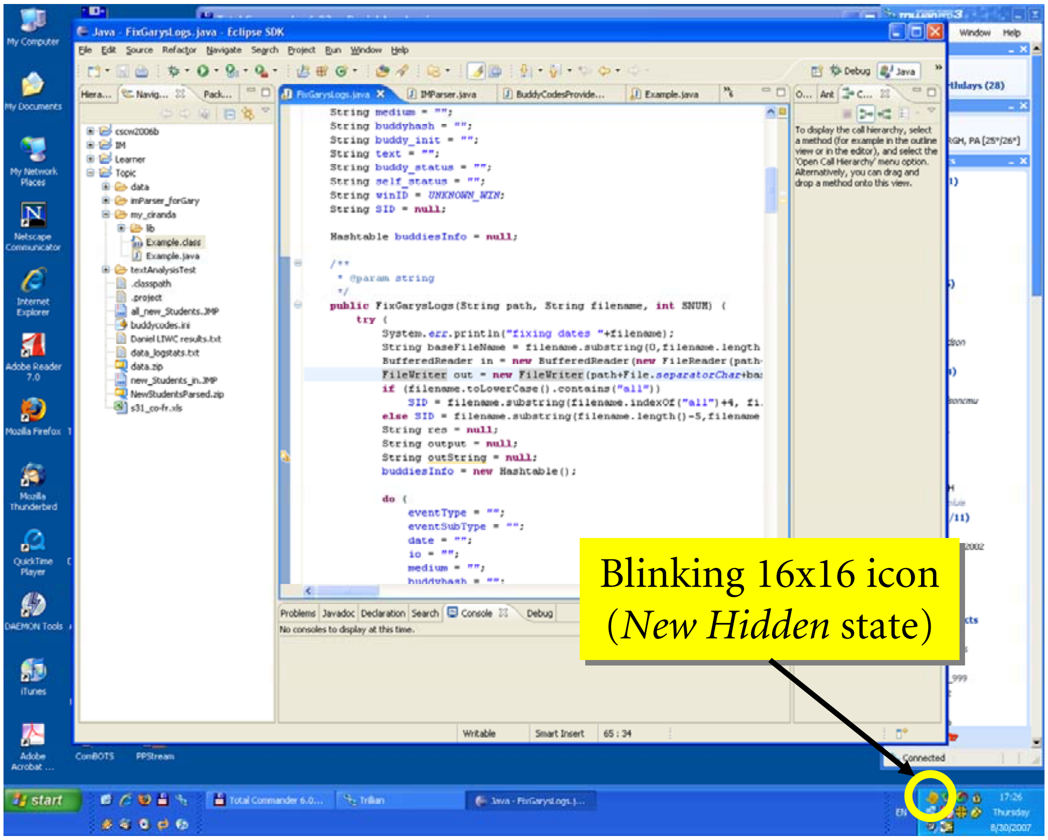


Figure 5.4 Notification of a new message in the *New Hidden* window state through a small (16x16 pixels) blinking icon.

Note that, regardless of a user's preference for the behavior of new message windows, existing windows will behave in the same manner (specifically, the flashing taskbar icon for windows that are *Existing Not Focused*).

The reader will notice immediately that each of these five possible states of the message window (*Existing Focused*, *Existing Not Focused*, *New Popup*, *New Minimized*, and *New Hidden*) have different attributes that may affect responsiveness to incoming messages. As mentioned above, it is likely that a window that was already open suggests recent communication. We may thus expect responsiveness to communication in these windows (whether in focus or not) to be faster than responsiveness to new communication attempts (in particular in this semi-synchronous setting in which users may ignore or postpone attending to incoming communication). On the other hand, the visibility of the message window upon the arrival of the message is likely to affect responsiveness. One would expect the very salient appearance of a newly-created message window (*New Popup*) and also windows already open and in focus (*Existing Focused*) to result in faster responsiveness than to messages in windows that are "out of sight". Furthermore, messages in windows in the *New Hidden* state, *New Minimized* state, as well as *Existing Not Focused* message windows require additional user action in order to display the message (clicking on the taskbar icon, using the keyboard to bring the message to the foreground, or clicking on the systray icon). This additional action may result in slower responsiveness.

H2: The state of a message window will have a significant effect on responsiveness.

H2a: An incoming message arriving when a message window was already open, whether in the foreground or background, will have faster responsiveness (since it is likely associated with ongoing communication) than messages appearing in a newly created window.

H2b: The high salience of the message window in the Existing Focused and the New Popup as well as the fewer user-actions required to attend to the message will result in significantly faster *responsiveness than for incoming messages that arrive in windows that are "out of sight" (Existing Not Focused, New Minimized, and New Hidden).*

One will note that if hypothesis 2a is correct, then *Existing Not Focused* should be associated with faster responsiveness than *New Popup*. Hypothesis 2b, on the other hand, suggests faster responsiveness in the *New Popup* state due to the high salience of messages in a *New Popup* state and/or the additional action required in the *Existing Not Focused* state.

In order to examine the effect of message window state on responsiveness and test hypotheses 2, 2a, and 2b, I performed a mixed model analysis where responsiveness (log transformed) was the dependent measure. The full set of measures described in Section 5.3 was included, with Window State (Existing Focused, Existing Not Focused, New Popup, New Minimized, and New Hidden) as the main independent measure of interest. ParticipantID was nested in Group and BuddyID was nested in ParticipantID then in Group. Both ParticipantID and BuddyID were modeled as random effects. (Remember that only incoming messages for which the participant's status was *Online* were used in the analysis.)

The analysis showed that the state of the window when an incoming message arrived has a large and significant effect on the user's responsiveness to the message ($F[4,31692]= 560$; $p<.001$; see Figure 5.5). This confirms hypothesis 2. A pair-wise comparison using Tukey HSD revealed a number of significant differences between the states. As expected, messages arriving in a window that is *Existing Focused* were responded to significantly faster (M=24 seconds) than in any of the other states. Next, responsiveness was fastest when a newly created window displayed as the focused application (*New Popup*; M=55 seconds) and significantly faster than the remaining three states. Messages in the *Existing Not Focused* (M=91 seconds) were significantly faster than both messages in the *New Minimized* state (M=123 seconds) and messages in the *New Hidden* state (M=156 seconds). The *New Minimized* and *New Hidden* states were not significantly different from one another.

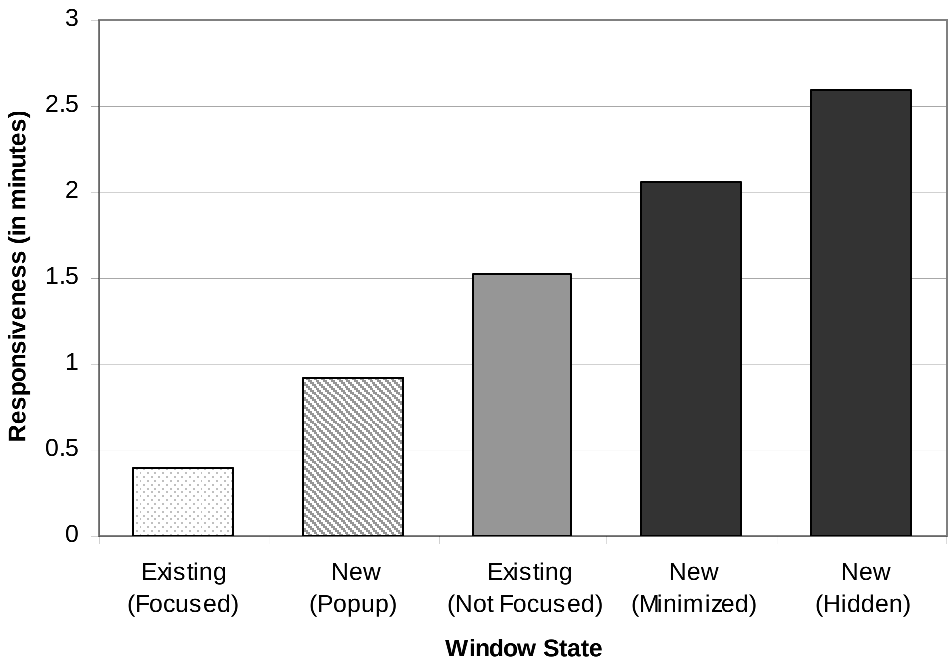


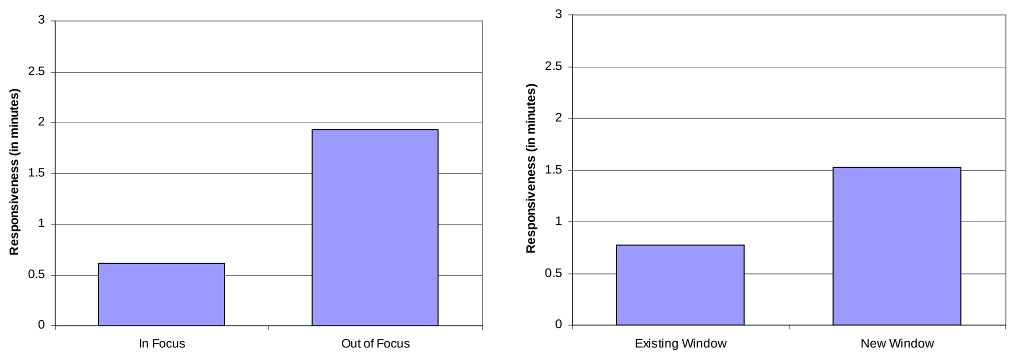
Figure 5.5 The effect of the state of the message window on responsiveness. Bars with different shades are significantly different from one another.

The significantly faster responsiveness for messages in a *New Popup* state compared to messages in windows that were *Existing Not Focused* suggest that the salience of an incoming message has a stronger effect on responsiveness than that of whether the communication was ongoing. Note that the significant difference between *Existing Focused* and *New Popup*, and the significant difference between *Existing Not Focused* and *New Minimized*, as both pairs share similar salience, suggest that whether communication was ongoing does in fact affect responsiveness, although the effect is smaller than that of salience.

The results of this analysis indicate that the analysis of responsiveness that follows next needs to take the state of the window into account and suggest that the effect of measures may be different in the various windows states. In order to address this potential interaction between measures and message window state, I created two new binary measures that represent the state of the window – Window (New vs. Existing) and Focus (In Focus vs. Out of Focus). For example, the *New Hidden* as well as the *New Minimized* window states were coded as Window(New) and Focus(Out of Focus).

Re-running the analysis described above, this time with Window(New vs. Existing) and Focus(In Focus vs. Out of Focus) and the interaction term (Window x Focus) showed a significant effect of whether the window was new or existing ($F[1,31922]=129.1$, $p<.001$; see Figure 5.6a). This confirms hypothesis 2a. The analysis shows an even larger effect of the salience of a window (whether in focus or not) on responsiveness ($F[1,31993]=536.3$, $p<.001$; see Figure 5.6b) confirming hypothesis 2b.

The significant interaction of Window and Focus ($F[1,31235]=12.0$, $p<.001$) is presented in Figure 5.7. Similar to the findings from the previous analysis, the salience of the message window appears to have a stronger effect on responsiveness than whether the window was new or existed already.



(a) **(b)**

Figure 5.6 The effect of Window (a) and Focus (b) on responsiveness.

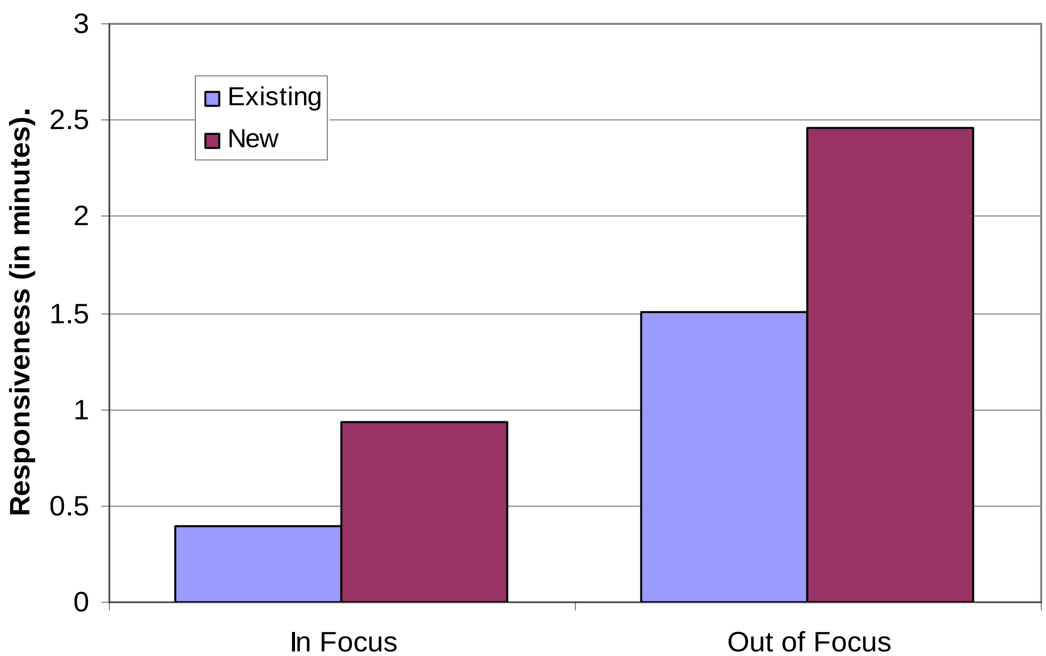


Figure 5.7 The effect of the interaction between Window (New vs. Existing) and Focus (In Focus vs. Out of Focus) on responsiveness.

5.6 The Main Analysis

The initial findings presented above influenced the final analysis in two ways. First, the effect of the user's online-status on responsiveness led to the decision to exclude from the final analysis data for which the user either explicitly indicated unavailability or was indicated to be inactive, leaving only data for which the user was *Online*. Second, the strong effect of the salience of the message window (whether the window was in focus or out of focus) on responsiveness led to the choice to examine, separately, the effect of the salience of the message window (using the Focus measure) and the effect of the window created or having existed (using the Window measure).

This main analysis was done as a mixed model analysis. Responsiveness, or the time until a response is sent (log-transformed) was the dependent measure. The full set of context, communication, and control measures listed in Section 5.3 were included as independent measures. The state of the message window measure was replaced with the Window (Existing vs. New) and the Focus (In Focus vs. Out of Focus) measures and the 2-way interaction between them Window*Focus. I also included the following 2-way interactions MultipleIMWindows*Window, MultipleIMWindows*Focus.

ParticipantID and BuddyID were modeled as random effects. Further, since each participant belonged to only one participation group, Participants were nested in Group. Similarly, since buddies appeared, for the most part, on only a single participant's buddy list, BuddyID was nested first in ParticipantID, then in Group. This analysis allowed controlling for differences in communication characteristics that originate from the differences between the participation groups or that originate from individual (or dyadic) differences.

5.7 Results

The analysis found a large number of significant effects on responsiveness (the results are summarized in Table 5.1). I will describe the results in each of the different measures categories.

5.7.1 Global Context

As expected, Day of Week had a significant effect on responsiveness ($F[6,28167]=10.6$, $p<.001$) and so did the Part of Day ($F[3,27639]=12.1, p<.001$; see Figure 5.8). In my data, responsiveness was significantly faster during the morning hours (M=64 seconds) and at night (M=65 seconds) compared to responsiveness during lunch and evening (M=76 and 77 seconds respectively; $t(26953)=5.855$, $p<.001$).

5.7.2 IM Context

A user's IM context had significant effect on responsiveness. The length of time (log-transformed) that the user was in an online status had a significant effect ($F[1,25720]=98.4$, $p<.001$) with quicker responsiveness when the user hadn't been online for long. Similarly, the time (log-transformed) since the user sent a message to a *different* buddy had a significant effect on responsiveness (F[1,30596]=10.1, p<.001). Responsiveness was faster when communication with others was more recent.

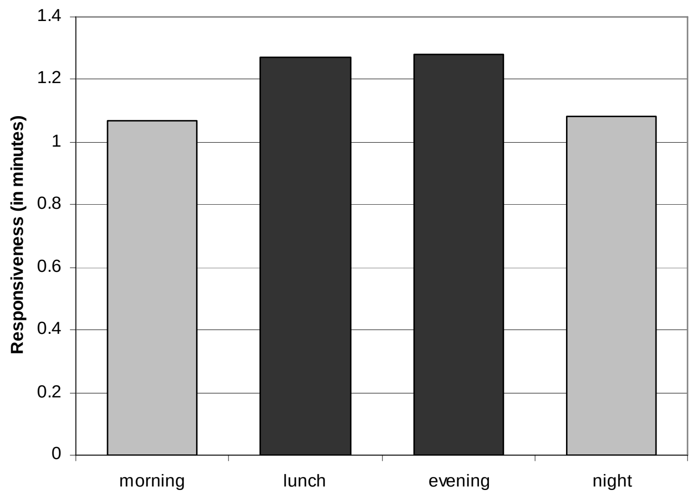


Figure 5.8 The effect of the part of the day on responsiveness. Bars with different shades are significantly different from one another.

The analysis found no main effect of the presence (or absence) of other IM windows on responsiveness (F[1,32000]=.53, n.s.). There was, however, significant interaction between the presence of other IM windows and Focus (In Focus vs. Out of Focus) (F[1,31960]=17.8, p<.001; see Figure 5.9) and between the presence of other IM windows and Window (New vs. Existing) (F[1,31927]=29.4, p<.001). A planned comparison showed that the presence of other message windows had a significant effect on responsiveness when an incoming message arrived in a window that was out of focus. Messages received a slower response when other IM windows were present than not (M=127 vs. M=109; t(32002)=3.08, p<.005). However, when the message arrived in a window that was already in focus, responsiveness was much faster and the presence of other IM windows did not show a significant effect (M=40 vs. M=44 seconds; t(32041)=1.21, n.s.). This finding is interesting and stresses the significant role of the salience of an incoming message on the user's responsiveness.

Similarly, the presence of other IM windows did not show significant effect on responsiveness when the message arrived in a new window, however it did show a significant difference when the window already existed.

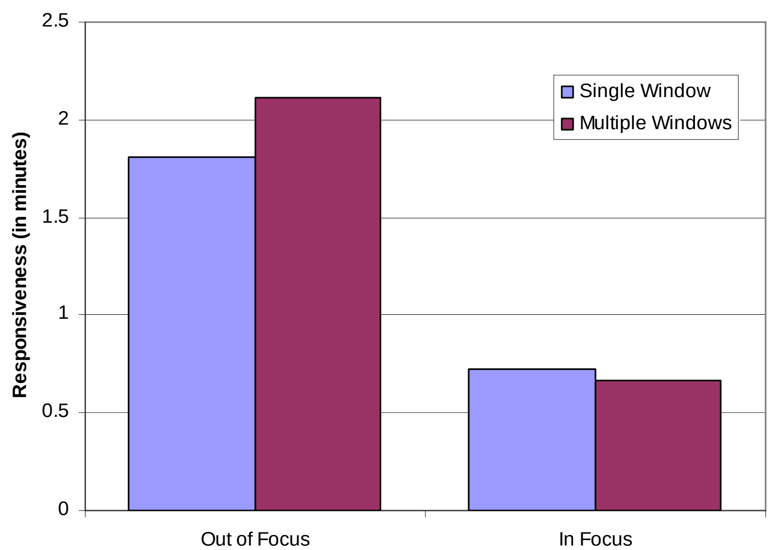


Figure 5.9 The effect of the interaction between Other IM Windows (Single vs. Multiple) and Focus (In Focus vs. Out of Focus) on responsiveness.

5.7.3 Desktop Context

The analysis showed that a user's ongoing activity on the computer had a significant correlation with their level of responsiveness to incoming messages.

The main application that was used on the computer in the two minutes prior to the arrival of a message had significant effect on responsiveness to that message ($F[21,31985]=4.6$, $p<.001$). The results show, for example, that using primarily a development tool (such as Microsoft Visual Studio, or the Eclipse programming environment) or a word processor, was correlated with significantly slower responsiveness to incoming messages. While also considered a productivity tool, the use of a statistics tool (such as SPSS, SAS, or JMP) was associated with significantly faster responsiveness. This oddity has at least two possible explanations. First, it is possible that participants were using IM to discuss the statistical analysis they were conducting. Second, it is possible that they were responding to incoming messages while waiting for the statistics tool to finish processing.

The results also show that the amount of keyboard activity prior to the arrival of the message had a significant effect on responsiveness ($F[1,31517]=102.8$, $p<.001$), as did the distance traveled by the mouse pointer ($F[1,31918]=40.2$, $p<.001$). The amount of window-title switches had a marginally significant effect on responsiveness ($F[1,31464]=3.1$, $p=.077$). In all three cases, greater *work-fragmentation*, in other words, an increase in activity (i.e., longer mouse movements, more keyboard activity, or more title switches) was correlated with faster responsiveness. One should keep in mind that these levels of computer activity were *prior* to the arrival of the message, not after its arrival.

Table 5.1 Results of a mixed model analysis of Responsiveness.
A negative (positive) estimate indicate that responsiveness is faster (slower).

Responsiveness (log-transformed) N = 32109, Mean Response = 1.5 (32secs)				
		Analysis of Variance		
Independent Variables	Estimate	*F*	d.f.	*p*
Global Context				
Day of the Week	*	6/28167	10.60	<.001
Part of the Day	*	3/27639	12.08	<.001
IM Context				
log(Time in Online Status)	0.088	1/25720	98.45	<.001
log(time since outgoing to a different buddy)	0.019	1/30596	10.06	<.005
MultipleWindows [multiple]	0.008	1/32000	0.53	---
MultipleWindows * Window	*	1/31927	29.36	<.001
MultipleWindows * Focus	*	1/31960	17.83	<.001
Desktop Context				
Type of App most used in last 2mins	*	21/31985	4.60	<.001
log(Keyboard Activity) – PCA	-0.047	1/31517	102.84	<.001
log(Mouse Distance) – PCA	-0.017	1/31918	40.21	<.001
log(Window-Title Switches) – PCA	-0.001	1/31464	3.13	.08
Communication				
Relationship Type	*	9/148	0.98	---
log(time since incoming from buddy)	-0.140	1/32041	539.51	<.001
log(time since outgoing to buddy)	-0.019	1/32045	6.99	<.01
Window [existing]	-0.175	1/31883	165.12	<.001
Focus [in focus]	-0.221	1/32023	376.05	<.001
Window * Focus	*	1/31612	28.17	<.001
Content				
log(length of msg in characters)	-0.196	1/32033	279.65	<.001
Is the msg a question? [yes]	-0.102	1/31753	189.16	<.001
Does the msg contain a URL? [yes]	0.166	1/32006	37.80	<.001
Does the msg contain an emoticon? [yes]	0.021	1/31878	4.05	<.05
Control Measures				
Participation Group	*	3/12	3.88	<.05
Age	-0.013	1/13	1.05	---
Gender [female]	-0.143	1/12	8.10	<.05
Lag-1	0.380	1/32046	4318.44	<.001

* Estimates for nominal measure with 3 or more levels are not in the table, rather discussed in the text.

5.7.4 *The Communication*

In general, elements of the communication in which the incoming message belonged had significant effect on responsiveness. As already discussed earlier, the state of the message window prior to the arrival of the message had significant and large effect on responsiveness. The time since the previous message sent to the buddy had significant effect on responsiveness ($F[1,32045]=6.99$, $p<.01$) as did the time since the previous message received from the buddy ($F[1,32041]=539.5$, $p<.001$). In both cases, longer time since the previous message was associated with faster responsiveness (a negative estimated coefficient).

Surprisingly, the type of Relationship did not have a significant effect on responsiveness ($F[9,148]=0.98$, n.s.). However, significant differences were found in responsiveness to different individuals (shown through predictions of the random effect of BuddyID and confirmed through a statistically significant increase in adjusted r-square with the inclusion of BuddyID as a random effect in the model).

5.7.5 *Content*

As expected, elements of the content of the messages itself played significant roles in responsiveness to the message. Messages that contained questions were responded to significantly faster (M=55 seconds) than messages that did not (M=89 seconds; $F[1,31753]=189.2$, $p<.001$; see Figure 5.10a). As predicted, messages that contained a URL received significantly slower responsiveness (M=103 seconds) compared to messages that did not contain a URL (M=48 seconds; F[1,32006]=37.8, p<.001; see Figure 5.10b). Messages that contained an emoticon were responded to significantly slower (M=74 seconds) than messages that did not (M=67 seconds; F[1,31878]=4.1, p<.05; see Figure 5.10c). Finally, the length of the message had a significant effect on responsiveness (F[1,32033]=279.6, p<.001) with faster responsiveness to longer messages (to be exact, the time to respond becomes shorter by 36% with every 10-fold increase in the length of the message).

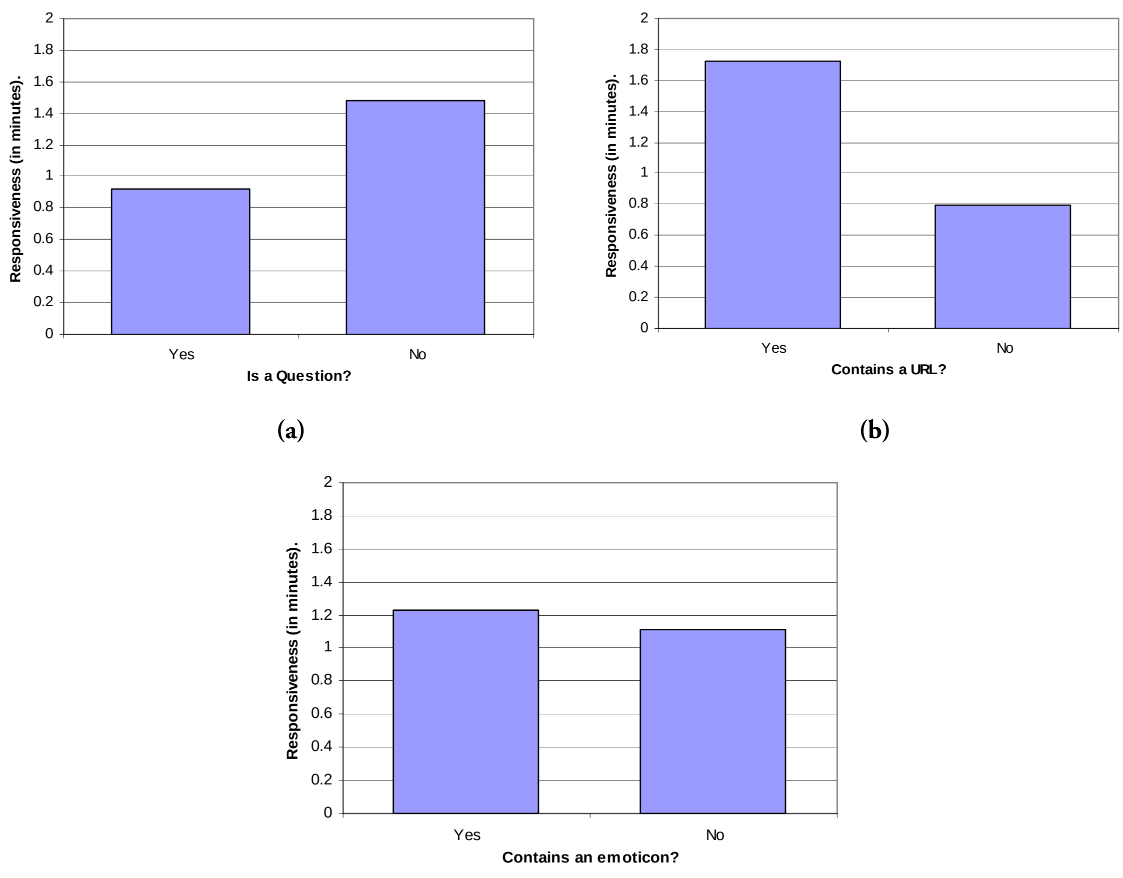


Figure 5.10 Content and Responsiveness. The effects of the presence of (a) a question, (b) a URL, and (c) an emoticon on responsiveness.

5.7.6 Control Measures

The analysis showed a couple of significant effects of the control measures on responsiveness. The participation group had a significant effect ($F[3,12]=3.9$, $p<.05$). A pair-wise comparison showed the participants in the Students group to have significantly faster responsiveness ($M=36$ seconds) than participants in the Startup group ($M=136$ seconds; $t(13)=3.3$, $p<.01$). Neither the responsiveness of the Students nor the Startup participants was significantly different from that of participants in the Interns group ($M=53$ seconds) nor the Researchers group ($M=93$ seconds). While the age of the participant did not show a significant effect on responsiveness ($F[1,13]=1.1$, n.s.), gender did have a significant effect on responsiveness, with the females responding to messages significantly faster, on average than the males ($M=50$ vs. $M=98$ seconds; $F[1,12]=8.1$; $p<.05$). Finally, the lag-1, or the responsiveness to the previous incoming message from the same buddy, had a large and significant effect on responsiveness ($F[1,32046]=4318$, $p<.001$).

5.8 Discussion

In the previous section I presented results from an in-depth analysis of factors that affect users' responsiveness to incoming instant messages. This analysis was performed on data collected in an unobtrusive fashion and without user intervention from participants' computers over extended periods. The findings show that many, although not all, of the measures examined had significant effect on responsiveness, confirming and expanding previous findings from laboratory experiments on communication coordination (Dabbish, 2007). These findings also provide further evidence of the link between the explicit behavior of responsiveness and the implicit state of availability. Finally, by observing communication at the beginnings, ends, as well as during conversations, this work is able to enhance our understanding of responsiveness, expanding previous research that examined responsiveness – or a user's willingness to engage in communication – only at the beginning of communication (see, for example, Dabbish & Kraut, 2004; Avrahami et al., 2007b).

5.8.1 The Effect of Computer Activity and Work-Fragmentation

I have presented evidence that users' ongoing computer activities prior to the arrival of a message significantly affect their responsiveness. This finding is in agreement with previous work on the interaction between subjects' primary task and their performance (and choices) when attending to an interrupting secondary task. In the real-world, incoming messages are none other than such interrupting secondary tasks (unless, of course, the IM communication was itself the user's primary activity). The analysis also showed a significant effect of the type of application used by participants on their responsiveness (for example, the slower responsiveness when using a programming environment). This finding is consistent with findings presented by Fogarty, Hudson and Lai (2004a). They found that features that described the computer applications recently used by their subjects were significant predictors of the subjects' self-reported interruptibility.

One of the interesting findings of my analysis is the significant inverse correlation between responsiveness and the user's *work-fragmentation* reflected by amounts of mouse activity and frequency of application switching. In other words, when users switch between applications frequently and display increased levels of mouse movement *prior* to the arrival of a message, they are likely to respond faster to incoming messages. (Naturally, once an incoming message has arrived, the user is likely to switch between applications making long mouse movements in the process.)

This finding suggests that users who are engaged in a task or tasks that involve frequent switching between applications are more receptive to incoming communication. Borrowing the terminology by Gonzales and Mark (2004), it is possible that users who are in-between work spheres are more willing to engage in additional external tasks such as incoming communication. It is important to point out that this finding is by no means obvious. It would have been quite plausible to expect high work-fragmentation to result in users taking longer to respond to incoming messages. Certainly, a high level of work-fragmentation can be an indication that the user is already juggling a lot of information and attending to another distraction may be undesirable. However, we hypothesize that infrequent switching between applications is associated with a user devoting their attention to a single task, resulting in unwillingness to be interrupted and in slower responsiveness.

5.8.2 The Effect of Content

The results show that measures related to the content of messages had significant effect on responsiveness. As discussed by Avrahami and Hudson (2004), messages containing questions are associated with the expectation of faster responsiveness as the asker is likely to be waiting for a reply. Not surprisingly then, incoming messages that contain questions in the data were responded to significantly faster. In contrast, messages that contained a URL received, on average, slower responsiveness. This is likely to result from messages that contain a URL requiring the receiver to follow that URL to some website before responding. The analysis also showed an interesting relationship between the presence of an emoticon in a message and slower responsiveness. The use of emoticons in IM allows users to express emotion or to clarify the tone of a message to avoid confusion. Oftentimes users will send messages that contain only an emoticon and no other text, indicating to the communication partner that the meaning of their message(s) was correctly understood. In those cases, the emoticon replaces non-verbal acknowledgements possible in other mediums. A quick examination of the relationship between the presence of an emoticon in an incoming message and the length of the message (controlling for the sender of the message) showed that messages were significantly shorter when containing an emoticon ($M=14$ vs. $M=20$ characters; $F[1,37433]=191.6$, $p<.001$). With these acknowledgements, the sender of the emoticon does not typically assume the floor, rather leaves the floor open (just as when providing non-verbal feedback in face-to-face conversations). The response to a message that contains an emoticon may thus be slower since it requires the user to initiate the next turn of conversation. Another explanation for this slower responsiveness to messages containing emoticons is that the use of emoticons in an incoming message may indicate a more relaxed style of conversation. I should note that these two possible explanations for the slower responsiveness in messages that contain emoticons are not mutually exclusive. Investigating the effect of other linguistic features on responsiveness, similar to that presented by Burke et al. (2007), would be valuable.

5.8.3 Influencing Responsiveness

Finally, I have presented earlier in this chapter a result showing the strong effect of the salience of the message window on responsiveness. The salience of the message window appeared to have a bigger impact on responsiveness than whether or not the message arrived in a new window or a window that has already existed. This finding is interesting, first, since it suggests that salience may have greater impact than whether a conversation was already ongoing. Second, this finding is interesting since it suggests that responsiveness could be programmatically influenced through changes in the method of delivery and presentation. This may allow the creation of enhanced communication applications that take advantage of knowledge of context and content (for example, through the use of predictive models similar to those presented in Chapter 3) in order to ensure that a user's attention is given to appropriate communication. I should note that this idea is by no means unique to IM. In mobile communication, for example, one might change the ringer settings on the phone automatically in order to influence responsiveness to incoming calls.

5.9 Conclusion

In this chapter I described results from an in-depth analysis of factors that affect responsiveness to incoming instant messages and discussed the link between responsiveness and availability. While this work describes investigation of responsiveness in a single medium (IM), the general classes of measures that were investigated – context, communication, and content – are not at all unique to IM, but generalize to other forms of interpersonal communication. While I collected data from individuals of different backgrounds and organizations, in this work, I did not examine the effects of these differences (rather controlled for them). It is thus still necessary to examine the impact of culture and norms on both demonstrated and desired availability in order to better understand the true relationship between responsiveness and availability under different settings. I propose also that it would thus be beneficial to investigate responsiveness as it is manifested in other media (and as different media interact). It would further be interesting to examine the change in the effect of measures of context and content when an incoming message is machine generated (as in a system message) and no longer part of interpersonal communication.

Chapter Six

Relationships and Communication Patterns*

In this chapter I report an investigation of the effect of interpersonal relationship on basic characteristics of IM communication (such as duration of session, length of messages, and the rate at which messages are exchanged), independent of message content. I then report on the use of findings from the analysis to inform the creation of two statistical models that classify the relationship between a user and their buddy based solely on basic communication characteristics (This work is described in Avrahami & Hudson, 2006a).

* The work presented in this chapter was originally published in Avrahami, D., & Hudson, S. E. (2006). Communication Characteristics of Instant Messaging: Effects and Predictions of Interpersonal Relationship. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW 2006)*, pp. 505-514. ACM Press.

6.1 Background

Our relationships with others affect our interaction with them in many ways (and this interaction, in turn, affects our relationship with them). Our relationship will affect the things we talk about, our style of communication and our perception of the value of the communication to us, our partner, and our relationship.

Duck et al. describe the effect of interpersonal relationships on everyday communication (Duck, Turr, Hurst, & Strejc, 1991). Using diary reports, they collected accounts of everyday spoken communication (either face-to-face or telephone) from over 1,700 students. Their analyses showed that interpersonal relationship type had significant effects on different aspects of communication, including the quality, purpose and perceived value of the communication.

Goldsmith and Baxter (1996) offer a taxonomy of communication styles (which they call "Speech Events"), such as formal, informal, involved, gossip, goal oriented, etc. They then show how different relationships are associated with different communication styles.

The growing popularity of electronic communication, such as email, IM, and Short Message System (SMS), raises similar interesting questions as to whether different relationship types would result in differences in electronic communication.

Feldstein (1982) and Crown (1982) describe the importance of cues such as tempo, pauses, speech rates and the frequency of turns, to the way in which participants in a conversation perceive each other. In this work I am interested in similar low-level aspects of communication – IM communication in particular – and how they are affected by the relationship between buddies.

Previous research has also shown significant differences in communication resulting from the frequency in which partners communicate and the frequency with which a communication medium is used by an individual and by the pair. Such differences were demonstrated in face-to-face communication (Whittaker, Frohlich, & Daly-Jones, 1994) and IM communication (Isaacs et al., 2002).

As more and more people use IM for their social as well as their work-related communication, I wanted to investigate the effect of interpersonal relationships on basic characteristics of IM communication (such as duration of session, length of messages, and the rate at which messages are exchanged). While interpersonal relationship might affect the use of grammar, abbreviations, or even the need to apologize for typos, in this work I examine its effect on more basic characteristics of IM, independent of message content, by answering the following two research questions:

- What, if any, are the effects of interpersonal relationship on basic characteristics of IM communication? And,

- if such effects exist, can basic communication characteristics be used to automatically classify the interpersonal relationship between a user and their buddy?

Automatic classification of interpersonal relationships are not interesting simply as a computational challenge but can in fact have real use in many different applications (e.g., an IM client that alerts users to incoming messages differently based on the classified relationship).

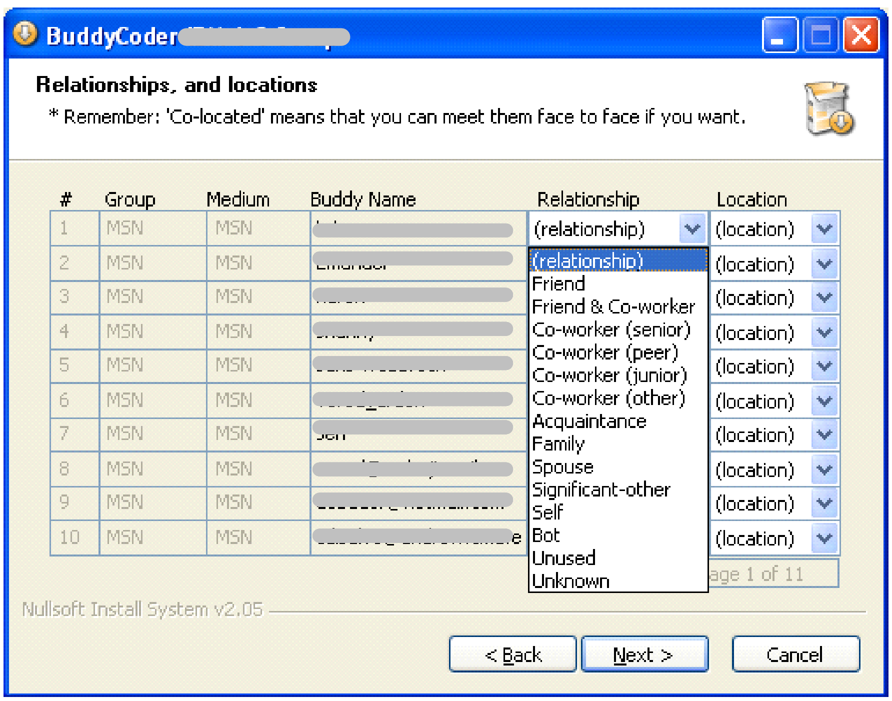


Figure 6.1 Application used by participants to classify buddies according to interpersonal relationships.

6.2 Data Collection

The data described in Section 3.3 was used for the analysis of the effects of interpersonal relationships. In order to obtain a classification of the relationships between participants and buddies, towards the end of their participation, each participant was asked to use a small coding program (see Figure 6.1) to indicate their relationship with each buddy in their buddy-list using the following 12 possible relationships: Co-worker (senior), Co-worker (peer), Co-worker (junior), Co-worker (other), Friend & Co-worker, Acquaintance, Friend, Family, Significant-other, Spouse, Self, and Bot. (A Bot is a computer program that users can communicate with through IM.)

6.3 Measures

6.3.1 Relationship Categories

As described earlier, at the end of their participation, each participant used a small coding program to indicate their relationship with each buddy in their buddy-list. For the analysis, the relationships were grouped into the following three higher-level relationship categories: Co-worker (senior), Co-worker (peer), Co-worker (junior), and Co-worker (other) were categorized as *Work*. Friend, Family, Significant-Other, and Spouse were categorized as *Social*. Friend & Co-worker was categorized as *Mix* and so was Acquaintance. Since my main interest was in interpersonal communication, the relationships classified as Self and Bot were excluded from further analysis.

6.3.2 IM Sessions

As in the previous chapter , the data were segmented into sessions by categorizing two instant messages as belonging to the same IM session if they were exchanged between a participant and their buddy within 5 minutes of one another (following Isaacs et al., 2002).

6.3.3 Communication Measures

For each IM session, a set of 12 measures was computed, describing basic characteristics of the session. These measures are:

- **Duration**: The length of time between the first and last message in the session (in minutes).
- **Message count**: The total number of messages exchanged in the session.
- **Turn count**: The total number of turns taken in the session. A single turn consists of consecutive messages sent by the same user.
- **Character count**: The total number of characters exchanged in the session (including spaces).
- **Messages-per-Minute**: The average number of messages sent per minute (Message count divided by Duration).
- **Messages-per-Turn**: The average number of messages sent per turn (Message count divided by Turn count).
- **Characters-per-Message**: The average length of messages (Character count divided by Message count).

- **Seconds Until First Reply**: The time between the end of the first turn and the beginning of the second turn (in seconds)*.
- **Minimum Gap**: The shortest gap between turns in the session (in seconds)*.
- **Maximum Gap**: The longest gap between turns in the session (in seconds)*.
- **Average Gap**: The average gap between turns in the session (in seconds)*.
- **Time of Day**: The time of the last message in the session.

To illustrate how these measures are computed, Table 6.1 shows the values of each of these measures computed for the transcript presented in Figure 2.2. For example, in this particular session the gap of 24 seconds between messages 4 and 5 represents the longest gap between turns in this session, named the Maximum Gap. The ratio of Messages-per-Turn is 12 / 7 = 1.71, and the average message length (Characters-per-Message) is 232 / 12 = 19.3.

6.4 Data Overview

To examine the effect of interpersonal relationship on basic communication characteristics I used the data set presented in Section 3.3 (and summarized in Table 3.1). The distribution of relationships as indicated by the participants is presented in Table 6.2 and Figure 6.2.

* Note that the value of this variable cannot exceed 5 minutes, since a gap longer than 5 minutes would qualify as the end of the session.

Table 6.1 Session measures computed for the session presented in Figure 2.2.

Variable	Value	
Group	Student	
Relationship	Work	
Duration	1.88	minutes
Message Count	12	
Turn Count	7	
Character Count	232	
Messages per Minute	6.4	
Messages per Turn	1.71	
Characters per Message	19.3	
Seconds Until First Reply	1	seconds
Minimum Gap (between turns)	1	seconds
Maximum Gap (between turns)	24	seconds
Average Gap (between turns)	12.2	seconds
Time of Day	5:44	pm

6.4.1 Excluding Single-Turn Sessions

Single-turn sessions are IM sessions in which one user sends one or more messages without a reply. 1190 of the total sessions in the data were identified as single-turn sessions. (A large number of those represent failed communication attempts.) Since single-turn sessions provide very little information about the interaction between a participant and a buddy, we removed these sessions from all the analyses and modeling presented next. After excluding the single-turn sessions, the data set contained a total of 3297 sessions between 412 participant-buddy pairs.

Table 6.2 Distribution of Buddies by Relationship and Group.

(Note: A buddy appearing several times in a participant's buddy-list will also appear those many times in the data)

		Researchers	Interns	Students
Work	Co-worker (senior)	22	6	1
	Co-worker (peer)	43	6	24
	Co-worker (junior)	34	-	2
	Co-worker (other)	9	-	-
Mix	Friend & Co-worker	16	13	80
	Acquaintance	-	2	12
Social	Friend	4	22	98
	Family	1	5	20
	Significant-other	-	3	2
	Spouse	-	2	-
Other	Self	1	1	5
	Bot	-	1	-

6.4.2 Relationship Distribution

The distribution of relationships as indicated by the participants is presented in Table 6.2. We can see that some relationships appeared very little or were not reported at all by different participation groups. For example, the Researchers indicated 22 of their buddies as being in the Co-worker (senior) category, while only one buddy was identified in that category from the Students group. Figure 6.2 shows the proportion of each high-level relationship category as indicated by each participation group. (Note that if a buddy appears on a participant's buddy-list more than once using different buddy-names, then that buddy will also be counted more than once in the data.)

From both Table 6.2 and Figure 6.2, it is clear that the distribution of relationships is different between the participation groups. For example, 83% of the buddies that the Researchers communicated with were identified as Work, compared to 11% for the Students group. Similarly, over 49% of the buddies in the Students and Interns groups were identified as Social, compared to only 4% for the Researchers. These differences between the participation groups were controlled for in the analysis.

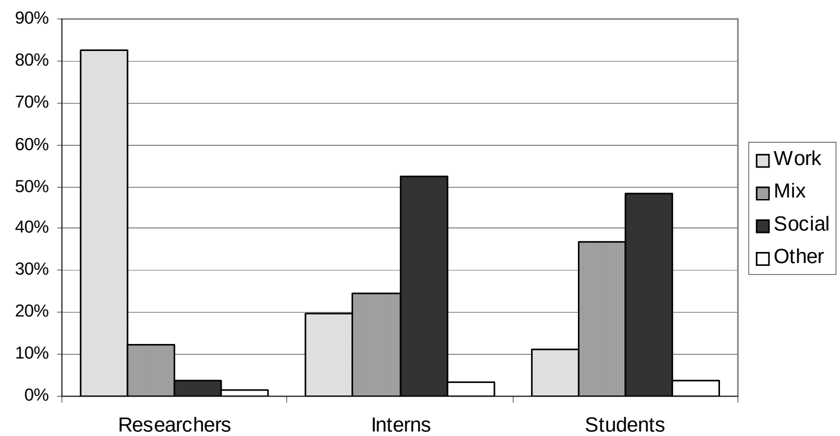


Figure 6.2 Distribution of Buddies by Relationship Category and Group.

6.5 Results

Table 6.3 shows the correlation coefficients for each pair of measures. As could be expected, the correlation between Duration, Message count, Turn count, and Character count is extremely high (r≥.88). It is also interesting to note that the inverse correlation between Messages-per-Minute and Average Gap is only r=-.25. The two are inversely correlated since, when message rate is higher, the gap between turns is likely to be shorter (recall, however, that message rate is related not only to gaps between turns, but also to gaps within turns).

To examine the effect of relationship on each of the communication characteristics variables described above, we used a mixed model analysis in which Relationship Category (Work, Mix, Social) and Group (Researchers, Interns, Students) were set as a fixed effect. Because participants and buddies typically communicated with one another more than once, observations were not independent of one another. Participants and BuddyID were modeled as random effects. Further, since each participant belonged to only one participation group, Participants were nested in Group. Similarly, since buddies appeared, for the most part, on only a single participant's buddy list, BuddyID was nested first in Participants, then in Group. This analysis allowed us to control for differences in communication characteristics that originate from the differences between the participation groups (evident in Table 6.2 and Figure 6.2) or that originate from individual (or dyadic) differences.

The results, summarized in Table 6.4, show that many of the communication characteristics were affected by the Relationship between the users and their buddies. Sessions of buddies in a Work relationship were shorter in duration – due in part to a smaller number of messages exchanged and to an overall faster exchange, although the length of messages themselves was longer. Here are the results in detail.

I found that Relationship had significant effect on Duration ($F [2,331] = 8.04$, $p<.001$). Sessions between buddies in a Social relationship lasted, on average, 2 and a half minutes longer (M=6.6 minutes) than sessions between buddies in a Work relationship (M=4 minutes) and about one and a half minutes longer than sessions between buddies in a Mix relationship (M=5.2 minutes)[*]. A planned pair-wise comparison showed that Duration of session was significantly different between sessions with buddies in a Social relationship and sessions with buddies in either Work or Mix relationships[†] (t(310)=3.65, p<.001, and t(331)=2.72, p=.007, respectively).

Since Duration, Message count, Turn count, and Character count are all correlated at over .85 (see Table 6.3) one could expect similar differences for these variables too. This is indeed true for the pair-wise comparisons between Social and Work relationships (Message count M=25.9 vs. M=13.8; t(382)=3.27, p=.001; Turn count M=15.3 vs. M=8.8; t(350)=3.28, p=.001; and Character count M=844.6 vs. M=459.5; t(316)=2.95, p<.004) but not for the Mix relationship.

I found that Relationship had significant effect on Messages-per-Minute – the pace with which messages were exchanged – (F [2,99] = 4.75, p=.01). Interestingly, we discovered that, while users tended to have longer sessions with buddies in a Social relationship and exchanged more messages per session, they exchanged messages with these buddies at a

---

[*] Because the independent variables were not completely orthogonal, Least Squared Means (LS Means) were used to control for the values of the other independent variables. The means reported throughout this chapter are LS Means.

[†] All pair-wise comparisons were done using the Tukey HSD post-hoc test.

significantly slower pace. Messages-per-Minute was significantly lower for buddies in a Social relationship compared to Mix relationship ($M=4.6$ vs. $M=6.2$ messages per minute; $t(115)=-2.99$, $p=.003$) and marginally significant compared to Work relationships ($M=4.6$ vs. $M=6.0$ messages per minute; $t(70)=-1.8$, $p=.078$). Messages-per-Minute did not vary significantly between Work and Mix.

A potentially related result is the significant effect of Relationship on Maximum Gap ($F[2,173] = 3.25$. $p<.05$), where a significantly longer maximum gap between turns was "allowed" in sessions with Social buddies ($M=82$ seconds) compared to sessions with Work buddies ($M=69$ seconds; $t(172)=-2.51$, $p=.013$). It is possible that the difference in Maximum Gap simply results from the fact that longer gaps are more likely in longer sessions that contain more turns. The correlation of $r=.46$ between Maximum Gap and the overall Duration of the session suggests that this explanation can account for a large portion of this effect but might not account for it entirely.

The results also show that Relationship had a significant effect on Characters-per-Message ($F[2,229] = 7.85$, $p<.001$). The length of messages exchanged between buddies in a Work relationship were longer, on average, than messages exchanged between buddies in either a Mix or a Social relationship ($M=38$ vs. $M=32$ or $M=30$; $t(1,219)=3.95$, $p<.001$ and $t(1,250)=3.11$, $p=.002$). Message length did not vary significantly between Mix and Social relationships.

We did not find significant effects of Relationship on any of the remaining communication characteristics variables. We did, however, find two significant effects of Participation Group on communication characteristics.

Participation Group had a significant effect on the average number of messages per turn (Messages-per-Turn) ($F [2,16] = 7.82$, $p<.01$), with the Students exchanging significantly more messages per turn than the Researchers ($M=1.7$ vs. $M=1.4$; $t(22)=-3.63$, $p<.002$). Messages-per-Turn was not significantly different between Interns and Students nor Interns and Researchers. This result is similar to results reported by Isaacs et al. (2002) where message exchange rate between their Light and Heavy IM users differed significantly (in their work, they used the term "turn" to refer to what we consider a single message). Paraphrasing their terminology, underlying differences between the participation groups, and in particular the Researchers and Students could warrant classifying them as Heavy and Super-Heavy respectively (See Table 3.1).

Participation Group also had significant effect on Time of Day ($F [2,15] = 36.8$, $p<.001$). This is not surprising considering that unlike the Students, the Researchers and Interns used IM primarily during business hours. This result is in accordance with results found by Begole et al. (2004).

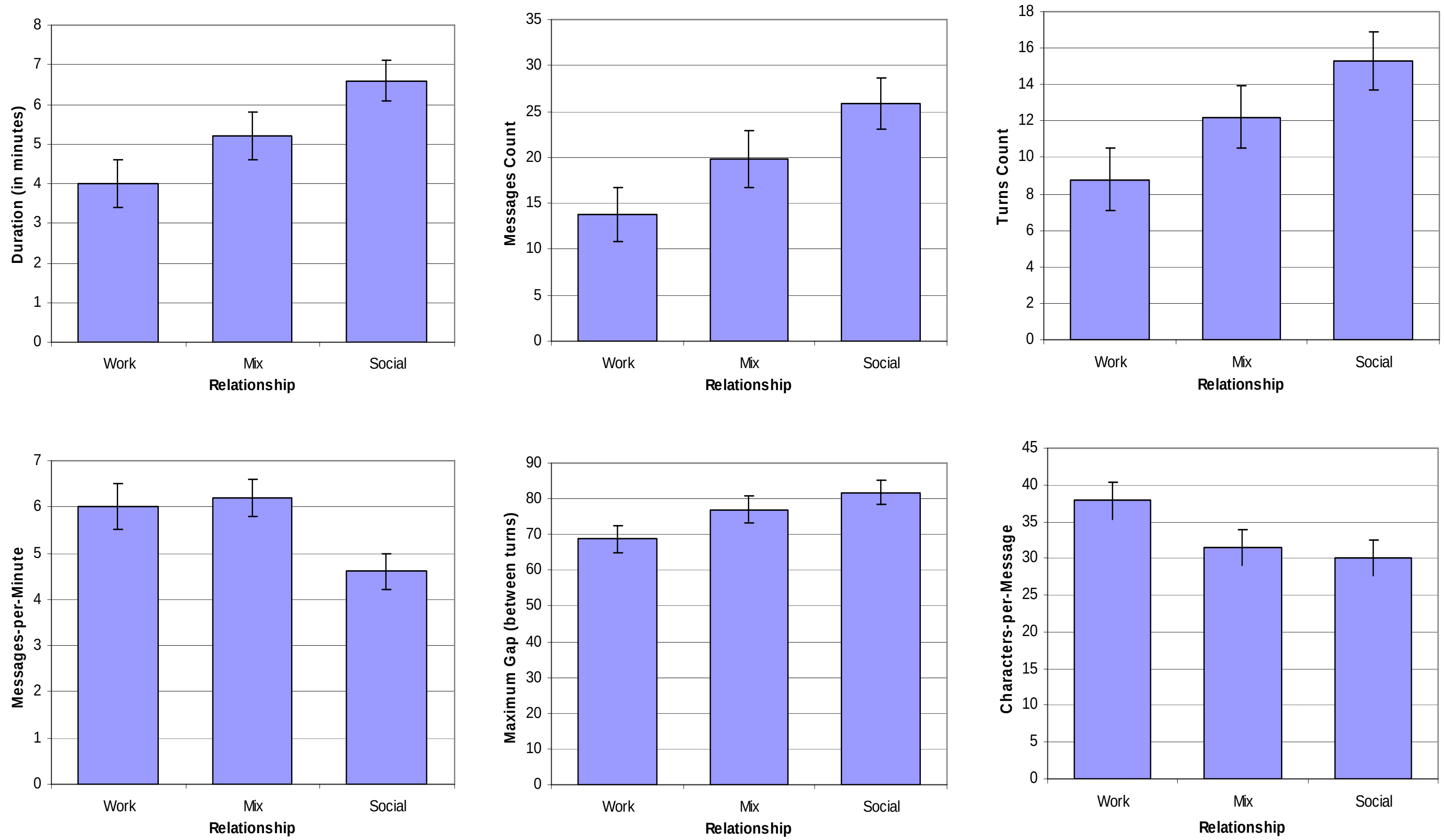


Figure 6.3 Effect of Relationship on Duration, Messages Count, Turns Count, Messages-per-Minute, Maximum Gap, and Characters-per-Message

Table 6.3 Correlation coefficients of the IM characteristics variables (N=3297)

	Duration	Message Count	Turns Count	Chars Count	Message per Minute	Message per Turn	Chars per Message	Secs Until First Reply	Minimum Gap	Maximum Gap	Average Gap
Message Count	0.88										
Turns Count	0.88	0.99									
Characters Count	0.88	0.95	0.95								
Message per Minute	-0.15	-0.04	-0.04	-0.05							
Message per Turn	0.15	0.16	0.07	0.12	-0.06						
Chars per Message	0.11	0.03	0.05	0.18	0.01	-0.06					
Seconds Until First Reply	0.03	-0.08	-0.09	-0.08	-0.20	0.09	0.04				
Minimum Gap	-0.12	-0.16	-0.17	-0.15	-0.12	0.07	0.05	0.56			
Maximum Gap	0.46	0.22	0.22	0.22	-0.30	0.12	0.07	0.49	0.25		
Average Gap	-0.01	-0.17	-0.18	-0.14	-0.25	0.09	0.08	0.69	0.87	0.58	
Time of Day	0.17	0.17	0.16	0.12	0.01	0.16	-0.11	0.00	0.00	0.08	-0.01

Table 6.4 The effect of relationship on IM characteristics (N=3297)

	Relationship Category								
	Work		Mix		Social		Analysis of Variance		
Variables	Mean	StdErr	Mean	StdErr	Mean	StdErr	*F*	d.f.	*p*
Duration (in minutes)	4.0	0.6	5.2	0.6	6.6	0.5	8.04	2/331	<.001
Messages Count	13.8	3.0	19.8	3.1	25.9	2.8	6.11	2/398	<.01
Turns Count	8.8	1.7	12.2	1.7	15.3	1.6	5.96	2/374	<.01
Characters Count	459.5	122.7	673.6	123.6	844.6	115.2	4.71	2/340	<.01
Messages-per-Minute	6.0	0.5	6.2	0.4	4.6	0.4	4.75	2/99	<.05
Messages-per-Turn §	1.5	0.05	1.5	0.05	1.6	0.05	2.32	2/312	
Characters-per-Message	37.9	2.5	31.5	2.5	30.1	2.4	7.85	2/229	<.001
Seconds Until First Reply	36.9	3.0	35.0	3.1	36.0	2.7	0.11	2/151	
Minimum Gap (between turns)	12.0	1.8	12.4	1.9	12.1	1.6	0.02	2/111	
Maximum Gap (between turns)	68.7	3.8	77.0	3.9	81.8	3.4	3.25	2/173	<.05
Average Gap (between turns)	28.8	2.2	28.3	2.3	29.2	2.0	0.10	2/181	
Time of Day §	14.6	0.4	14.6	0.4	14.7	0.4	0.04	2/253	

§ - Participation Group (Researchers, Interns, and Students) had significant effect on this variable

Table 6.5 The effect of multiple ongoing IM communication on IM characteristics (N=3297)

	MultipleSessions				Analysis of Variance		
	0 (Not Engaged)		1 (Engaged)				
Variables	Mean	StdErr	Mean	StdErr	*F*	d.f.	*p*
Duration (in minutes) *	4.5	0.4	6.6	0.5	44.66	1/3188	<.001
Messages Count *	16.9	2.2	24.6	2.3	27.26	1/3227	<.001
Turns Count *	10.4	1.3	14.9	1.4	29.35	1/3243	<.001
Characters Count *	564.9	99.4	828.6	103.4	25.22	1/3264	<.001
Messages-per-Minute *	6.2	0.3	4.8	0.3	13.93	1/471	<.001
Messages-per-Turn §	1.5	0.04	1.5	0.04	0.66	1/3238	
Characters-per-Message *	33.4	2.2	32.7	2.3	0.92	1/3222	
Seconds Until First Reply	36.9	2.2	34.2	2.4	2.00	1/2953	
Minimum Gap (between turns)	12.9	1.4	10.8	1.5	2.96	1/2852	
Maximum Gap (between turns) *	79.1	2.6	83.2	2.9	19.93	1/2739	<.001
Average Gap (between turns)	28.6	1.6	29.0	1.7	0.06	1/3080	
Time of Day §	8.5	0.3	9.0	0.3	10.12	1/3255	<.002

* - Relationship category (Work, Mix, Social) had significant effect on this variable

§ - Participation Group (Researchers, Interns, and Students) had significant effect on this variable

6.6 Discussion

The analysis described above showed the significant effect of relationship on a number of the communication characteristics investigated. It was no surprise to find the effect of relationship on the overall length of sessions (including duration, number of messages, turns, and characters). However, I was surprised and intrigued by the effect of relationship on message exchange rate (Messages-per-Minute) and on the average length of messages (Characters-per-Message).

One possible explanation for the interesting differences in message length is that conversation between buddies in a work relationship is less casual and users construct their ideas more carefully before sending them. Another explanation could be that conversation with work buddies requires greater verbosity to achieve common ground than conversation with social buddies. Finally, it is possible that the concepts discussed with work relationships (perhaps more complex) simply require the use of longer terms to describe.

Based on my findings, the difference in message exchange rate between Work, Mix, and Social relationships cannot simply be accounted for by differences in the length of messages. In fact, the results show the exact opposite. Not only did participants and buddies in a Work relationship exchange longer messages on average, but they also did so at a faster pace overall. An interesting possible explanation for the differences in pace is that users devoted different levels of their attention to the different conversations. In other words, it is possible that users focus less of their undivided attention to conversations with their Social buddies and give more attention to conversations with their Work buddies. This explanation is supported, in part, by the significant differences in the Maximum Gap between turns. The Maximum Gap reflects the maximum time that users let their conversation partners wait before responding. The significantly higher gap allowed between buddies with a relationship of a social nature may again suggest that less focus of attention is given to sessions with those buddies in comparison to conversation with buddies in a work relationship. Another possible explanation for the difference in message exchange rate is that communication with Social contacts happens when other communication is taking place, and it is presence of multiple ongoing conversations that accounts for the difference in message exchange rate. I examined this explanation more closely below.

6.6.1 Follow-up Analysis: Relationship, Parallel Communication, and Communication Patterns

One of the interesting findings presented in Section 0 revealed that relationship had a significant effect on the pace with which messages were exchanged, with buddies in a Social relationship exchanging messages, on average, at a significantly slower rate. In order to examine the possibility that this difference in message exchange rate is a result of users being simultaneously engaged in multiple IM sessions, I conducted two follow-up analyses.

The first analysis examined the effect of engagement in multiple simultaneous IM sessions on message exchange rate. The second analysis examined whether multiple simultaneous IM sessions was more or less common when communicating with buddies in different relationships and in each of the different participation groups.

6.6.1.1 Engagement in Simultaneous IM

A binary measure of a user's engagement in other IM sessions was computed for every session as follows: Every incoming or outgoing message in the session was marked as 1 (engaged) if a message to *another* buddy was sent or received within the last five minutes, and 0 (not engaged) otherwise. The session measure *MultipleSessions* (0 or 1), was then computed as a Boolean OR of this indicator for all messages in the session. That is, MultipleSessions equals 1 for a particular session if one or more of the messages in that session were indicated to have taken place simultaneously with another session, and 0 otherwise. Using this coarse measure of simultaneous engagement, 37% of sessions (1224) in the data were identified as taking place simultaneously with other sessions.

To investigate the effect of engagement in simultaneous IM on basic communication patterns and in order to examine whether such an effect will subsume the effect of Relationship found earlier, I repeated the mixed model analysis described in Section 0, this time with MultipleSessions (0 or 1) as an additional fixed effect. I will now briefly describe the results (the results are presented, in full, in Table 6.5).

In accordance with findings presented in Chapter 5, MultipleSessions indeed had a significant effect on many of the basic communication patterns. The length of sessions, and the number of messages, turns, and characters exchanged were all significantly affected by other ongoing IM communication (Duration, Message count, Turn count, and Character count). Message exchange rate was also significantly affected by other sessions (F [1,471]=13.9, p<.001). Messages were exchanged at a significantly slower pace when other IM communication was taking place (M=6.2 vs. M=4.8 messages-per-minute). The more important finding for this investigation, however, was that the effect of Relationship category was still present (F [2,106]=5.86). Message exchange rate was again shown to be significantly slower for sessions between buddies in a Social relationship (M=4.45) compared to buddies in a Mix relationship (M=6.18; t(138)=3.39, p<.001) and marginally significantly slower compared to sessions between buddies in a Work relationship (M=5.8; t(70)=1.79, p<.077).

To examine whether multiple ongoing IM sessions were more common in any of the participation groups or when communication between the different relationship categories took place, I conducted a mixed model logistic regression analysis, in which MultipleSessions was the dependent measure, Relationship category (Work, Mix, Social) and Group (Researchers, Interns, Students) were set as fixed effects. As in the analysis in Section 0 ParticipantID was nested in Group and set as a random effect, and BuddyID was nested in ParticipantID and then in Group and set as a random effect. This analysis found no significant effect of Relationship category or Participation Group on MutipleSessions.

From these analyses we may thus conclude that the effect of multiple ongoing IM communication cannot by itself account for the significant difference in communication patterns between the different Relationship categories.

6.7 Automatic Classification of Interpersonal Relationships

In this section I describe the creation of two predictive models (or "classifiers") that classify the relationship between a user and their buddies using only those basic characteristics shown in the previous section. Both models were generated using Nominal Logistic Regression. (Other classification techniques, including Naïve Bayes and Decision Trees were also explored but resulted in lower accuracy.) Both models used a similar two-step process to provide their classifications. In the first step, the model classifies the relationship for each individual IM session, and in the second, a majority vote is taken for each participant-buddy relationship, across all their joint sessions, to provide a final classification.

It is important to stress that the models attempt to classify the *relationship* between IM buddies, not the content of their individual conversations (although the two are undoubtedly related). That is, a model should classify friends as being in a Social relationship even if they sometimes talk about work. Similarly, a model should classify co-workers as being in a Work relationship even though they may discuss the location for an after work drink.

Automatically classifying the relationship between IM users can be used in a number of ways. First, such classifications could be used to augment IM systems. For example, a system such as Lilsys (Begole et al., 2004) could set indicators of unavailability to buddies individually, based on relationships. IM clients could also alert users to incoming messages differently, depending on their relationship with the sender. An augmented IM client that observes the content of incoming messages -- similar to the client described in Chapter 7 and in Avrahami and Hudson (2004) -- or a client that predicts whether a user is likely to respond to a message (using models such as the ones presented by Avrahami & Hudson, 2006b) could use classification of relationships to help guide whether or not to increase the salience of incoming messages. A completely different category of uses for these classifiers would be to allow the classifications, originating in IM, to propagate to other communication mediums. With many of today's IM service providers, such as Microsoft, AOL, Yahoo! and Google also providing email (and recently also Voice-over-IP), a person's IM identity (their buddy name) is often their email identity as well. Thus, an automatic classification of the relationship with a person, based on their IM interaction, could be used to enhance the interaction with the same person in different mediums. For example, such classifications could be used to inform systems such as the Priorities system that predict email interaction (Horvitz et al., 2002). Finally, automatic classifiers of relationships could also be used to provide an overview of IM communication in a whole organization, and even comparison between organizations.

The next section describes this process in detail followed by results and classification accuracy.

6.7.1 Preparing the Data

Informed by the results presented in the previous section, I used the following eight variables (or features) in the classifiers: Duration, Message count, Turn count, Character count, Messages-per-Minute, Messages-per-Turn, Characters-per-Message, and Maximum Gap. I could not use (or control for) Group or Participant in the models as these are not independent of relationship. In order for these models to be interesting, they had to work well across groups and without knowledge of the group that a participant belongs to (otherwise, if one knows, for example, that a participant belongs to the Researchers group then one could simply guess that the relationship with a buddy is a Work relationship and be correct 84% of the time). In order to make up for the inability to control for differences between the groups and participants, I applied a natural-log transformation to each of the variables (except for variables that represent rates). Thus, the final set of variables was as follows: log(Duration), log(Message count), log(Turn count), log(Character count), Messages-per-Minute, Messages-per-Turn, Characters-per-Message, and log(Maximum Gap).

6.7.2 Model 1: Work vs. Social

The first of the two models classifies relationships into one of two classes: Work or Social. For this model, I used a subset of the data containing only sessions between participants and buddies in either a Work or Social relationship. This subset contained 2379 sessions with

292 participant-buddy pairs (of which 203 pairs, or 70%, communicated in two sessions or more).

To test the accuracy of the model, I used a 16-fold cross validation method. That is, the model was created over 16 trials, one trial for each participant, and the combined accuracy is reported. Typically, with cross-validation, the data are randomly divided into a number of subsets. In this case, however, different sessions from the same participant are not independent (especially sessions with the same buddy), and randomly segmenting the data would likely result in some of a participants' sessions appearing in both the training and test data. This would give the model an unfair (and unrealistic) advantage. Instead, I used a more conservative cross-validation method in which, for each trial, the full data of a single participant is excluded as a test set and the data from the other participants are used for training.

6.7.2.1 Training Process

The training process for each trial follows three steps: First, all sessions of one participant are excluded and kept as a test set. Next, the remaining data are adjusted to contain an equal number of sessions for each class (described below). Finally, the model is generated using the sessions in the training set.

Adjusting the distribution of the training set is important in order to prevent the underlying bias in the distribution of sessions from biasing the classifications of relationships (for example, while only 37% of the buddies were identified by the participants as in a Social relationship, over 45% of sessions recorded were with those buddies). This bias in distribution was mostly a result of variance in the amount of data recorded from the different participants. Participants in the Researchers and Interns groups, for example, tended to use IM during business hours on weekdays, while participants in the Students group used IM nearly 24 hours a day, 7 days a week. Therefore, the data contain a greater number of sessions from the Students. Thus, prior to training a model, the training set is adjusted to include an equal number of sessions for each relationship category. This prevents the model from merely classifying relationships as Social as a result of their high frequency in the data. For example, if the training set consists of 700 Work sessions and 800 Social sessions, then 100 Social sessions are selected at random and excluded from the training set.

6.7.2.2 Classification Process

The classification performed by the models follows a two-step process (illustrated in Figure 6.4). First, the model is used to provide a relationship classification of 0 (Work) or 1 (Social) for each session in the test set (Figure 6.4a). We will refer to these classifications as "Session-level classifications". In the second step (Figure 6.4b), a single final classification is provided for each buddy using a majority vote among all session-level classifications for the same buddy. In other words, the model provides a final classification based on whether the average session-level classification is greater or smaller than 0.5. The second step is performed only for buddies with whom a participant had two or more sessions. In case of a tie (the average equals 0.5), the majority classification of all session-level classifications (for all buddies) is assigned as the final classification for the buddy. Figure 6.4b includes an illustration of a case where a tie is resolved (in this case, to generate an incorrect classification).

SN	Buddy	Actual	Session Variables	Prediction
i	*b1*	0 *(Work)*	$< v_1,v_2,\ldots,v_n >$	0 *(Work)*
i	*b1*	0 *(Work)*	$< v_1,v_2,\ldots,v_n >$	0 *(Work)*
i	*b1*	0 *(Work)*	$< v_1,v_2,\ldots,v_n >$	1 *(Social)*
i	*b2*	1 *(Social)*	$< v_1,v_2,\ldots,v_n >$	0 *(Work)*
i	*b2*	1 *(Social)*	$< v_1,v_2,\ldots,v_n >$	1 *(Social)*

(a) Step 1: Predict Relationship for each Session

SN	Buddy	*n*	Actual	Average	Final Prediction	Correct
i	*b1*	3	0 *(Work)*	.333	0 *(Work)*	Yes
i	*b2*	2	1 *(Social)*	.5	0 *(Work)*	No

(b) Step 2: Predict Relationship for each Buddy using average of individual Session predictions

Figure 6.4 Classification Process illustration: (a) Session-level classifications, and (b) final Buddy-level classifications with one correct and one incorrect classification.

6.7.2.3 Performance Results (Model 1)

The performance of this first model, for buddies with two or more sessions, is presented in Figure 6.5. The model was able to accurately classify 161 of the 203 relationships, for an accuracy of 79.3%; significantly better than the 53.2% prior probability (G2 (1,203)=73, p<.001). (Prior probability represents the accuracy of a model that picks the most frequent answer at all times.)

I was curious to see the model's performance when classifying relationships for buddies with whom the participants communicated only once. As expected, the accuracy of these classifications was much lower (41.6%). I believe that it is not unreasonable, however, for a system using such a model to require at least two data points before providing a final classification of relationship.

	Classified as: Work	Classified as: Social
Work	40.9% (83)	5.9% (12)
Social	14.8% (30)	38.4% (78)
	Accuracy: 79.3%	

Figure 6.5 Classification results of a model classifying Work vs. Social relationships.

	Classified as: Work	Classified as: Mix	Classified as: Social
Work	25.3% (74)	5.1% (15)	2.0% (6)
Mix	8.2% (24)	14.7% (43)	7.8% (23)
Social	9.6% (28)	17.1% (50)	10.2% (30)
	Overall Accuracy: 50.2% Work vs. Rest: 75.1% Social vs. Rest: 63.5%		

Figure 6.6 Classification results of a model classifying Work vs. Mix vs. Social relationships.

6.7.3 Model 2: Work, Mix, Social

Since the full data set consisted also of buddies with whom the participants were in a relationship that was a mix of both social and work, I next attempted the much harder 3-way classification problem. For this model I used the full data set, which contained 3297 sessions with 412 participant-buddy pairs (of which 293, or 71%, appeared in two or more sessions). Again, a 16-fold cross-validation was used, excluding the data from one participant each time, and training on the remaining data. The combined accuracy of the 16 trials is reported.

6.7.3.1 Training Process

The training process was almost identical to the process used for the 2-way model. In addition to adjusting the training set to contain an equal number of Work and Social sessions, training sets were adjusted to also include an equal number of Mix sessions.

6.7.3.2 Classification Process

Again, a two-step classification process is used, similar to the process described earlier. In the first step, the model provides a relationship classification of 0 (Work), 0.5 (Mix), or 1 (Social) for each session in the test set. In the second step, a single final classification is provided for each buddy using a slightly modified voting step among all session-level classifications for the same buddy. This time, final classifications are assigned as follows: If the average session-level classification for a buddy is greater than 2/3, then the model provides a final classification of 1 (Social). Similarly, if the average is less than 1/3, then the relationship is classified as 0 (Work). If the average is greater than 1/3 and smaller than 2/3, then the relationship I classified as 0.5 (Mix). In case that the average equals 1/3 or 2/3, a slightly more complicated process is used to resolve the tie. If the average classification for all sessions (with all buddies) is greater than 0.5, then an average session-level classification of 1/3 or 2/3 results in final classifications of 0.5 (Mix) or 1 (Social) respectively. Conversely, if the overall average for all sessions (of all buddies) is smaller than 0.5, then an average session-level classifications of 1/3 and 2/3 (for a single buddy) result in final classifications of 0 (Work) and 0.5 (Mix) respectively.

6.7.3.3 Performance Results (Model 2)

The performance of this second model is presented in Figure 6.6. The model was able to accurately classify 147 of the 293 relationships. This model's accuracy was only 50.2% (compared to the prior probability of 36.9%). Again, the accuracy of classifications for buddies with whom the participants communicated only once was even lower (36.1%). A closer examination of the model's classifications shows that the model was much more accurate at distinguishing Work from not Work (75.1%) than it was at distinguishing Social from not Social (63.5%).

6.8 Discussion

The performance of the first model (classifying Work vs. Social) was surprisingly high (nearly 80%) considering that no content of messages was used to generate the classifications. The drop in accuracy when moving to the 3-way model (classifying Work vs. Mix vs. Social) could be a result of the greater difficulty of a 3-way classification in general. However, I believe that the main reason for this drop in accuracy is that the Mix relationship is, indeed, similar to both the Work and Social relationships. I am examining the possibility of using a cascading approach, in which a model first classifies whether a relationship is Work or not, then a second model attempts to distinguish Mix from Social.

Indeed it is possible that the features used by the models are simply insufficient for distinguishing between all three of the relationship categories. This may suggest that different features are needed in order to accurately distinguish between the three categories, and in particular distinguish Mix from Social. These features may need to use some aspects of the content of messages (for example, using the Linguistic Inquiry and Word Count program (Pennebaker, Francis, & Booth, 2001)). Still, these models present an exciting potential for classifying relationships without using the private and potentially sensitive content of messages.

6.9 Conclusions and Future Work

In this chapter I described an analysis of the effect of interpersonal relationship on basic characteristics of IM communication. I presented, for example, a number of results that suggest that, while IM sessions with social contacts are longer in duration, users focus, on average, less of their undivided attention to these sessions. These findings add to previous research, which showed the effect of interpersonal relationships on face-to-face and phone communication, by extending it to IM communication. This work also complements previous research that described the effect of frequency of communication on basic characteristics of communication in both synchronous and asynchronous mediums.

I used the results of the analysis to inform the creation of two models that automatically classify interpersonal relationships based solely on basic characteristics of communication. One of the models described was able to classify, with 79.3% accuracy, whether a user and a buddy are in a work or social relationship. This accuracy is impressively high considering that only basic characteristics of communication were used, without knowledge of the actual content of messages. Finally, I discussed the results and potential uses for automatic classifiers of interpersonal relationship.

Using a sample of 16 participants meant that the data set, while not small, contained conversations between only 412 participant-buddy pairs. Still, I believe that the findings should generalize beyond the 412 pairs in the set. Specifically, the relatively high performance of the first classifier, despite the significant differences between the participation groups (in age, profession, composition of buddy-list, etc.), suggests a robustness of the underlying findings.

In the work presented in this chapter, I grouped the fine-grain relationship categories presented in Table 6.2 into three high-level categories (Work, Mix, and Social). This grouping was done, in part, due to the uneven distribution of fine-grain relationships in the data. In a future data collection phase, I plan to expand the list of relationships to also include types shown by previous literature as having distinct properties (such as Best Friend). I will then examine, in detail, the effect of fine grain relationship categories on communication (e.g., do communication characteristics differ between sessions with a peer and with a senior co-worker?). However, it is important to remember that, from a machine-learning perspective, attempting to classify closely related concepts can be very difficult. As the performance of the models dropped with the introduction of the Mix relationship, one can expect a classification of all 10 fine-grain relationships to be very difficult. Goldsmith and Baxter (1996) proposed that relationships may need to be discussed not only in sociological terms (e.g., co-worker, friend) but also in ways that reflect the native construction of relating (e.g., "We have the kind of relationship in which you can tell everything," "I have a 'joking around' relationship"). Obtaining a new classification in such terms from participants will be required in order to re-examine the effect relationships on communication from this perspective.

Kraut et al. showed that physical distance has significant effect on coordination and communication (Kraut et al., 1990). I am interested in examining whether and how physical distance between IM buddies affects their basic communication characteristics. I plan to use the scale from Cummings and Ghosh (Cummings & Ghosh, under review) to get a coding of distance from future participants. I suspect that interesting differences exist in the interaction of relationship and physical distance.

CHAPTER SEVEN

Balancing Performance and Responsiveness[10]

7.1 Introduction

While many of the benefits of IM come from its near-synchronous nature, it is the asynchrony that allows users to multitask. With computers often permanently connected to the internet, users are able to keep their IM clients running continuously in the background. This means that incoming messages often arrive when the user is engaged with other tasks, possibly in the midst of intensive work. As O'Conaill & Frohlich (1995) and Kraut and Attewell (1997) point out, it is often the case that time and topic are convenient for the initiator (in this case, the buddy) but not the recipient. Results from a study conducted by Avrahami et al. showed that by merely assigning the role of initiator or recipient in a role-

---

[10] The work presented in this chapter was originally published in Avrahami, D., & Hudson, S. E. (2004). QnA: Augmenting an Instant Messaging Client to Balance User Responsiveness and Performance. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW 2004)*, pp. 515-518. ACM Press.

playing study (assigned to be either "Callers" or "Receivers"), one can observe a significant imbalance between initiators' frequent choice to initiate communication, and recipients' less frequent desire to accept it (Avrahami et al., 2007b).

In an attempt to alleviate the problem of IM disrupting work on an important task, or users being forced to ignore incoming messages in order to maintain workflow, I have created a tool, called QnA, for automatically alerting users to specific messages that may deserve their attention – in particular to potential questions and answers. (This work is described in Avrahami & Hudson, 2004)

7.2 Background

An instant message is regarded as a less intrusive way of interrupting than a phone call or a visit. IM further offers users "plausible deniability" (Nardi et al., 2000), that is, the ability to deny presence or receipt of a message, even after having read it. However, the common alerts associated with incoming messages (the message window opening, sound, and flashing or bouncing icons), even if brief, can easily distract the user and interfere with their work (the effects of interruptions on performance was discussed earlier in this document).

Being disrupted by message alerts is made worse by the fact that most IM clients have identical alerts for all incoming messages, not taking into account the identity of the sender or the content of the message. In addition, users will often send many short messages in succession even when these constitute a single conversational turn. Isaacs et al. suggest that experienced IM users are more prone towards this behavior (2002). Other research further suggests that this behavior may be influenced by elements of the IM message window (Gergle, Millen, Kraut, & Fussell, 2004). The result is the user being subjected to a large number of identical alerts, as many as one alert for each of these short incoming messages.

What then can a user do to handle the distractions from incoming messages?

Block all messages: Nardi et al. (2000) report that some users complained about being distracted by alerts while working towards important deadlines. These users reported having to resort to shutting IM down. One of the drawbacks to this type of strategy is that it relies on memory and appropriate planning by the user (having to remember to turn the IM client back on when they are available for communication). More importantly, this strategy ignores factors such as the identity of the sender and importance and urgency of the conversation, (mis-)treating urgent and non-urgent conversations equally. As Isaacs et al. note, most IM conversations held in the workplace are work-related (2002), which makes closing the IM client an undesirable strategy.

Set availability indicators: Another strategy available to IM users is to change their online status indicator. This option allows them to indicate to their buddies that they are busy or unavailable. This strategy too, however, has a number of drawbacks. First it depends on buddies recognizing and not ignoring these indicators. It also requires users to plan ahead and set their status appropriately. (Changing one's online status after receiving a message can be socially awkward since it directly signals to the initiator that they have been disruptive). In addition, this strategy runs the risk that users may forget to reset their status once they are available again, making these indicators unreliable (A possible solution to this last problem was presented by Begole et al. (2002) in a system that learns the user's work rhythms over time, providing buddies with estimates of the user's online presence).

Read all, respond to some: since the sender of an instant message cannot automatically know whether or when their message was read, users are able to read or skim incoming messages before choosing whether or not to respond. This is similar to the use of a caller ID in telephones, where users can determine the source of a call before selecting to accept it. The main benefit of this method in IM is that users may have some idea about the topic of conversation before fully engaging in it. However, this method can also become a burden as it requires users to devote a fair amount of their attention to messages merely in order to decide to ignore them.

Ignore until reaching a breakpoint: Finally, users can elect to stay on task and simply ignore, to the best of their ability, the alerts of incoming messages. With most IM clients, however, this strategy can be quite difficult. In particular, users are unable to determine which of the incoming messages could be ignored for some time and which require their immediate attention.

The solution described here allows users to employ this strategy while providing them with a mechanism for distinguishing between incoming messages.

7.2.1 Expectations for Responsiveness

As mentions above, different messages are associated with different expectations for levels of responsiveness. These range from messages for which a sender is expecting a quick response (e.g. – in the message "*do you have the figures I need for the meeting?*"), those for which a leisurely response is sufficient (e.g. – "*check this out www.interesting.com*"), messages that can be politely deferred (e.g. – "*busy?*"), to messages that do not need a response at all (e.g. – "*going to a meeting. ttyl*[11]"). Ignoring or delaying response to messages that are associated with expectations of a quick response may not only portray the user as impolite or even rude, but may also adversely affect the buddy if they need information to proceed with their work.

In order to allow users to take advantage of these differences in expectations for responsiveness, I have created a tool called *QnA* that helps users identify messages that potentially require a quick response (and messages that they are expecting), distinguishing them from messages that could potentially be ignored for some time. This tool allows users to stay on task, while appearing responsive to those buddies who are expecting quick

[11] ttyl is a common abbreviation for "talk to you later".

responses. More specifically, I chose to notify users on incoming questions, and incoming answers to their own questions.

Figure 7.1 QnA notifications:
(a) Question (b) Possible response (c) Question and possible response
(d) Notification with preview

The following scenario illustrates the use of this tool:

7.2.2 Illustration of Use

Jim is in his office preparing a presentation for a meeting that same afternoon. As usual, he is running an IM client with QnA in the background for fast communication with his colleagues. He is missing a few figures and sends an IM to his colleague Bill "did you mean to remove the figure from slide 5". Bill does not reply and Jim goes back to the presentation. Being short for time, Jim ignores a couple of incoming messages when he notices a QnA notification saying that Bill may be replying to his question (Figure 7.1b). Jim clicks on the notification, bringing the message from Bill to the front. It reads "no, definitely not". Jim then notices a QnA notification saying that Liz is asking him a question (Figure 7.1a). He clicks on the notification to find a message that reads "do we have a projector?" As Jim is typing his reply, Liz sends another question and Jim modifies his reply. Since Jim was typing, QnA determines it doesn't need to show a notification for Liz's second question.

7.2.3 Why Questions and Answers?

The choice to notify users on questions and answers results from the important role that the question and answer pair plays in human dialogue. In particular, it was noticed that a party in a conversation who asks a question will expect a response and is unlikely to disengage from the conversation (unless a response fails to arrive for some time). Schegloff and Sacks define the concept of adjacency pairs in conversation and give question-answer pairs as one type of adjacency pairs (Schegloff & Sacks, 1973):

> "…adjacency pairs consist of sequences which properly have the following features: (1) two utterance length, (2) adjacent positioning of component utterances, (3) different speakers producing each utterance." (p.295)

> "…a first pair part and a second pair part…form a 'pair type'. 'Question-answer', 'greeting-greeting', 'offer-acceptance/refusal', are instances of pair types. A given sequence will thus be composed of an utterance that is a first pair part produced by one speaker directly followed by the production by a different speaker of an utterance which is (a) a second pair part, and (b) is

> from the same pair type as the first utterance in the sequence is a member of." (p.296)

Clark, in his book "Arenas of Language Use", describes the question-answer pair as the prototype of adjacency pairs (Clark, 1992), stating:

> "Adjacency pairs consist of two ordered utterances, the first and second pair parts, produced by two different speakers. […] One crucial property is conditional relevance. Given a first pair part, a second pair part is conditionally relevant, that is, relevant and expectable, as the next utterance. Once A has asked the question, it is relevant and expectable for B to answer in the next turn." (p. 157)

We can regard an incoming instant message that contains a question to be representing a first pair part (thus a response from the user is "relevant and expectable") and an incoming instant message in response to a question is regarded as a second pair part (thus the user is likely to be expecting it). If it is established that the user did not attend to these messages for a certain period of time, QnA notifies the user of the pending message, the identity of the sender, and whether the message represents a question, a possible response to a question, or both.

7.3 Implementation

QnA was implemented as a plug-in for Trillian Pro and is available for download to Trillian Pro users from the Plugin Development forum on the Trillian website or from the author's homepage. It was written in C and implemented as a Dynamically-Linked-Library (DLL) that is run from inside Trillian Pro. Identifying that a message window received focus was done using the Windows CBTHook HCBT_SETFOCUS command.

7.3.1 Events and Flow-Control

QnA is composed of two main processes, presented in Figure 7.2. The first process monitors incoming and outgoing instant messages while the other monitors user actions on incoming messages. These processes are described in more details next.

QnA uses three internal flags for every buddy the user is sending or receiving messages from. These flags allow QnA to keep track of messages and to determine whether it should present a notification to the user. The flags are: `expectingResponse`, `incomingResponse`, and `incomingQuestion`.

7.3.1.1 Processing Outgoing Messages

When the user sends an outgoing message to a buddy, QnA scans the message and, using a set of string matching rules, determines whether or not the message is likely to contain a question (For description and discussion of the set of rules used see Section 7.3.3). If it estimates that the message contains a question, it then sets an internal flag called `expectingResponse`, indicating that the user may be expecting a response from this buddy. If QnA estimates that the message does not contain a question, it does nothing.

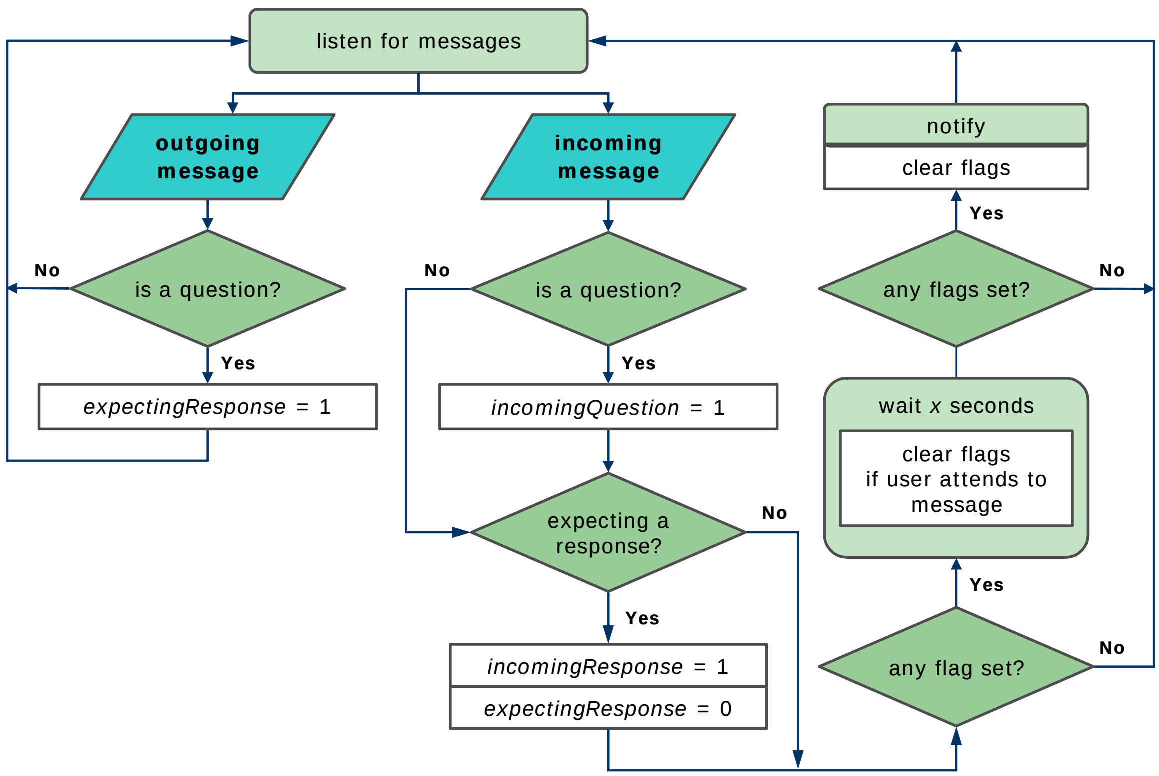


Figure 7.2 Flow control and internal flags of QnA

7.3.1.2 Processing Incoming Messages

When an incoming message from a buddy is received, QnA first estimates whether the message contains a question using the same string matching rules. If so, it sets an `incomingQuestion` flag, indicating internally that the user might want to respond to this message. It also checks whether or not the `expectingResponse` flag was set for this buddy. If it was, then it is reset, and the `incomingResponse` flag is set instead, indicating that the buddy may have responded to a question.

7.3.2 QnA Notifications

If either the `incomingQuestion` or `incomingResponse` flags is set (or if both are) then QnA initiates a process responsible for establishing whether the user is attending to the message. This process waits a certain number of seconds (configurable by the user, with a default value of 10 seconds). If, at the end of the wait period the `incomingQuestion` flag is set but not the `incomingResponse` flag, a small (non-modal) notification similar to the alert shown in Figure 7.1a is presented at the bottom right corner of the user's screen. If, however, only the `incomingResponse` flag is set, a notification similar to the one shown in Figure 7.1b is presented. If both flags are set the user is presented with an alert similar to the one shown in Figure 7.1c. Users are also able to configure QnA to replace the notifications described above with notifications that display a preview of the message, similar to the notification shown in Figure 7.1d, for any of the three cases described above.

After the notification is shown, all flags for the buddy are reset. This is done so that no more than one notification per conversation will be shown every wait period, allowing users to ignore the notifications more easily if they choose to.

Notifications automatically fade out and disappear after 10 seconds unless clicked on by the user. If clicked on, the notification disappears and the corresponding message window is opened (if the window is already open, it is brought to the foreground).

7.3.2.1 Suspending Notifications

Following the experience of using the very first version of QnA, I realized that the presentation of notifications about questions or answers from a buddy, while useful, should be suspended if the user is already engaged in conversation with that same buddy. Otherwise, we run the risk of constant interference with the already ongoing conversation. This was accomplished by introducing the delay period described earlier between the message arrival and the presentation of the notification. If during the delay the user types a message to a buddy, opens a message window for that buddy, or if the message window is in focus, it is assumed that the user will have seen any incoming message, and QnA notifications regarding messages from that buddy are suspended. This is done by resetting both the `incomingQuestion` and `incomingResponse` flags. This allows QnA to intercept notifications even if they are already in the wait period.

I specifically chose not to use closing of the message window as indicator of the user attending to the message since the user might close the window without realizing that a message has just arrived.

7.3.3 Identifying Questions

In order to determine whether a message contains a question, the message is compared against a set of string matching rules. QnA identifies the message as a question if any match is found. All matching performed is case-insensitive. The set of rules was adapted for typical

IM spelling and abbreviations. It was then further refined and expanded based on feedback from users (Figure 7.3 shows a partial list of the rules used).

```
'?' at the end of a line or sentence
'/' at the end of a line (a common typo for '?')
what (is|are|r|were|does|do|did|should|can)
where (is|are|r|were|does|do|did|should|can)
when (is|are|r|were|does|do|did|should|can)
how (is|are|r|were|does|do|did|should|can)
who (is|are|r|were|does|do|did|should|can)
did(|n’t|nt) (i|u|you|he|she|they|we)
do (i|u|you|he|she|they|we)
will (i|u|you|he|she|they|we)
should(|n’t|nt) (i|u|you|he|she|they|we)
(are|r) (you|u)
huh
```

Figure 7.3 String matching rules used to estimate whether a message contains a question (partial list)

A second set of rules was created to try and eliminate phrases that should not be considered to be questions for the purpose of notification, but that match at least one of the rules (these can be regarded as ‘false-positives’). These are mostly questions that serve the purpose of querying and negotiating the availability of the receiver and can be ignored (e.g., “are you there?”). One could argue that ignoring such questions serves, in a sense, as a response to them. Figure 7.4 shows a few of the rules used.

```
(are|r) (you|u) there
hello?
busy?
how (are|r) (you|u)
```

Figure 7.4 String matching rules for messages that should not be considered a question.

7.3.4 User Preferences

An important aspect of QnA is users' ability to customize the behavior of the plugin to fit their work and messaging style. (Figure 7.5 shows the user options dialog). Since the natural level of responsiveness is often different for different users, it is important that QnA allows enough time for the user to notice and attend to messages before displaying the notifications. Thus, the first and most important user-customizable option is the number of seconds that QnA waits before displaying a notification. (If set to zero, notifications appear instantaneously and no suspension of notifications can occur.) Users may select to be notified only on questions, or only on responses to their questions. Users can also decide whether notifications should be suspended when typing or when opening the message window. Suspending notifications if the message-window is in focus when the message arrives may be undesirable to some users, for example, in cases when IM is the primary application running and the user is not attending to the computer (e.g., while reading a paper document). Finally, users can choose whether notifications should show the typical QnA notifications (as in Figure 7.1 a, b, and c) or a preview of the message itself (as in

Figure 7.1d). In the future, users may also be able to personalize the list of string matching rules for identifying questions.

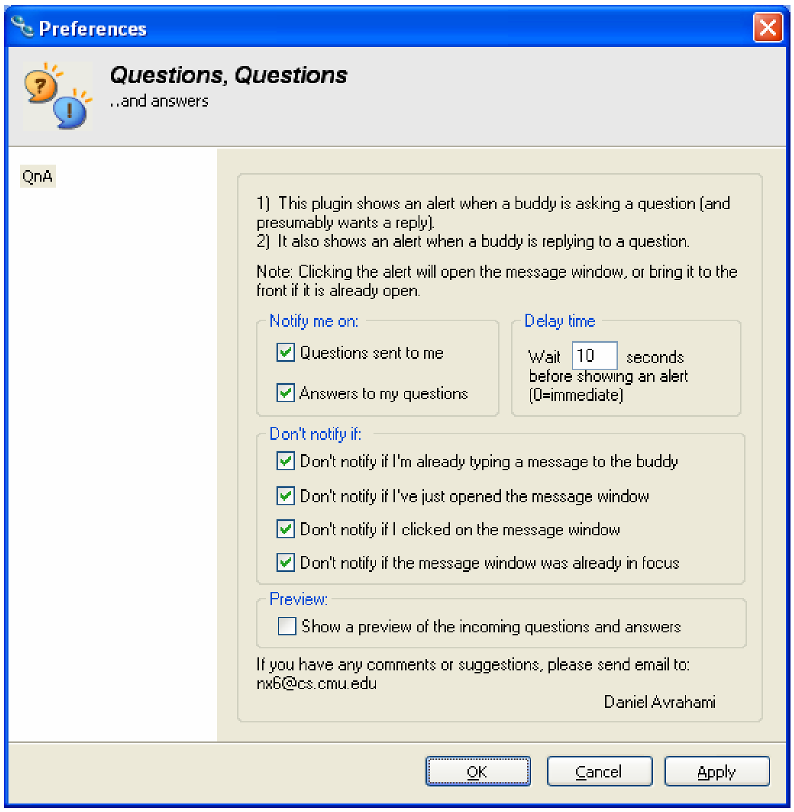


Figure 7.5 QnA user preferences. Users choose whether to be notified on incoming questions, answers, or both. Set their preferred delay period (showing with 10 seconds). Select events for which notifications will be suspended (typing, opening the window, etc.). Choose whether notification presents a preview of the incoming message.

7.4 Evaluation

In this section I describe results from a preliminary evaluation of the effect of QnA on user's IM interaction. This evaluation was done by analyzing the effect of the presence of a question in an incoming message on the time it took the user to open the message-window of that incoming message, and the effect of the presence of a question in an incoming message on the time until an already open, but out-of-focus window (*Open not Focused)*, was brought to the foreground by the user.

This analysis allows us to examine whether QnA had an effect on the time it took the participant to open the message window. Since the participant does not know the content of the message until the window is opened, a significant effect of the presence of a question in the message would indicate QnA's effect. Similarly, for windows that are open but are out of focus (*Open Not Focused)*, an effect of the presence of a question on the time until the message window is brought to the foreground can provide an indication of the effectiveness of QnA. (Note that this second indication is weaker since a message window that is out of focus can still be visible to the user).

The main hypotheses studied are as follows:

- H1a: Users will open message-windows of incoming messages that contain questions faster, on average, than windows of incoming messages that do not contain questions.
- H1b: Users will bring to the foreground message-windows that are already open but out of focus (*Open Not Focused*) faster, on average, when the incoming messages in those windows contain a question than when the incoming messages do not contain a question.

I will now describe briefly the data collected, the analyses performed, and the results found.

7.4.1 Data

For this preliminary evaluation I examined the use of QnA by one of the participants described in Section 3.3, who permitted the collection of the text of messages. The participant, belonging to the Startup participation group (See Table 3.1), recorded IM data for a period of five and a half months (158 days). During this period the participant communicated with 48 buddies, exchanging 32584 messages (17740 incoming and 14844 outgoing). The participant's preference for the delivery of messages when a window was not yet open was to be notified through a blinking icon at the bottom right corner of the screen. Finally, the participant used QnA throughout their participation period, and used preferences identical to those presented in Figure 7.5, with the exception of their choice of a five seconds wait period before notifications are shown (instead of default setting of a ten seconds wait period).

7.4.2 Evaluation Results

Two analyses were conducted to test hypotheses 1a and 1b:

To test hypothesis 1a, the first analysis used only those incoming messages for which the window was *Not Open* (n=4798). The time until the window was opened (log transformed) was the dependent measure. The presence of a question in the message (0 or 1) was the main independent measure of interest. Day of the Week (Mon – Sun) and Part of Day (Morning, Lunch, Evening, Night) were added as control measures. Since the participant communicated with buddies more than once, BuddyID was treated as a random effect.

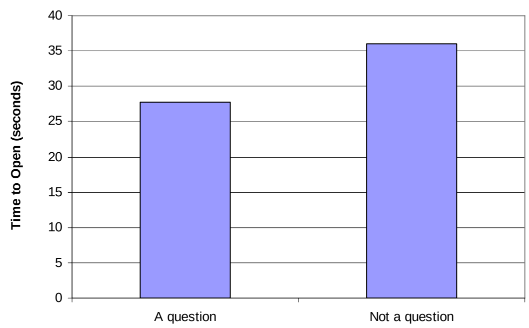
This analysis shows that the presence of a question in the message had a significant effect on the time until the window was opened ($F\ [1,4786]=14.59$, $p<.001$) with messages containing a question opened significantly faster (M=27seconds vs. M=36seconds; see Figure 7.6a). Significant differences were also found for the two control measures. Both Day of the Week ($F[6,4766]=3.11$, $p<.005$) and Part of Day ($F[3,4752]=6.73$, $p<.001$) were significantly correlated with the time to open the message window but could not account for the effect of the presence of a question.

The findings from this analysis thus support hypothesis 1a.

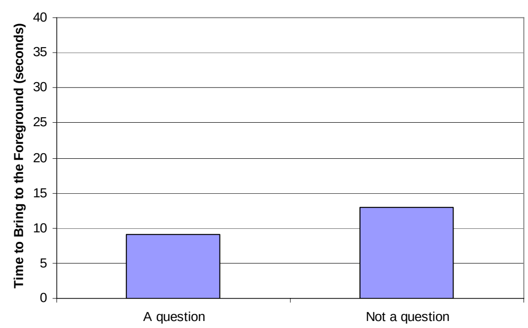
The second analysis, performed to test hypothesis 1b, used only those incoming messages for which the window was already *Open but Not Focused* (n=7900). The time until the window was brought into focus (log transformed) was the dependent measure. The presence of a question in the message (0 or 1) was the main independent measure of interest. Day of the Week (Mon – Sun) and Part of Day (Morning, Lunch, Evening, Night) were added as control measures. Since the participant communicated with buddies more than once, BuddyID was treated as a random effect.

Again, the presence of a question in the message had a significant effect on user actions, with the time until a window was brought into focus, significantly shorter when the message contained a question (M=13seconds vs. M=9seconds ; F[1,7887]=45.29, p<.001; see Figure 7.6b). Part of Day also had a significant effect (F[3,5508]=2.87, p<.05) and Day of the Week had a marginal effect (F[6,6915]=1.99, p=.064).

The findings from this second analysis thus support hypothesis 1b.



(a)



(b)

Figure 7.6 The significant effect of the presence of a question in an incoming message when using QnA on the time to: (a) open a message window that is not yet open and (b) bring a window that is open but out of focus to the foreground.

7.5 Discussion

7.5.1 Identifying Questions and Answers, and the Cost of Errors

Identifying questions and answers reliably in instant messages is a challenging task for a number of reasons. One such reason is that relaxed grammar and speling are the norm in IM (Nardi et al., 2000; Voida et al., 2002). Furthermore, instant messages often contain abbreviations. These include abbreviations for single words (for example, ‘u’ to mean ‘you’), or for whole sentences (for example, ‘ttyl’ to mean ‘talk to you later’). The message “r u ready 2 go”, for example, needs to be identified as a question. There are a number of reasons for this. The first is that IM buddies, as opposed to chat or email, are almost always familiar with one another. Users are less concerned about being perceived as ineloquent, giving priority to sending the message fast. The second reason, and possibly more important one, is the desire to keep the conversation as synchronous as possible. Delaying sending a message to correct spelling or fix grammar can slow the conversation down or even suggest a change in conversation turns. Thus, users may elect to send a message containing a grammatical or spelling error.

The mechanism used by QnA for identifying questions in instant messages in order to notify users of messages that may deserve their immediate attention is a simple one (specifically, the body messages is compared against a long list of string matching rules). But a simple mechanism of this sort cannot be error-proof. On one hand, messages that should not are identified as questions by QnA. For example, the message “and then he asked me: where are you going?” which is not intended as a question for the receiver. On the other hand, some messages that should be identified as question will be missed by QnA if they don’t match any of the rules in the set (although the set could potentially be expanded to reduce the likelihood of this happening). However, I claim that the cost associated with such occasional errors is low due to the interaction model employed by QnA. A message containing a question that is missed by QnA (a false-negative error) will still appear on the user screen as any other normal message. A QnA notification for a message that is identified as a question even though it was not intended as such by the sender will likely be viewed and quickly dismissed by the user. By designing QnA to augment a user's knowledge of different incoming messages rather than acting on those messages on the user's behalf, the cost of an occasional inaccuracy becomes low.

Identifying answers to questions reliably is also difficult. This is due primarily to the multi-threaded nature of IM conversations. As Voida et al. note, following a multi-threaded conversation can be so hard that it may even confuse the people participating in the conversation (2002). Researchers in the area of Natural Language and Information Retrieval are working hard to address the problem of identifying questions and matching answers (See for example, Agichtein, S., & Gravano, 2001; Zhang & Lee, 2003). The solutions they propose may indeed be useful for the tool described in this chapter. However, as the availability of message persistence can cause users to send many short messages (Gergle et al., 2004), an incoming message may in fact be part of an answer, but not the whole answer. This may prevent the more sophisticated solutions from providing significant improvement. I believe that notifying the user of the first incoming message following a question, combined with "cautious" notification wording ("*X* might be answering your question"), is a reasonable solution.

7.5.2 Misuse and Empowerment

Another issue worth discussing is that of the potential misuse of QnA. Specifically, QnA allows buddies who are aware of a user's use of QnA to increase the salience of their messages by simply adding, for example, a question mark at the end of each of their messages (the more subtle buddies may actually re-phrase their messages as questions). In some respects, this is similar to allowing the senders of email to associate a level of urgency with their email. There is, however, one major difference between email and IM that may alleviate this concern. While anyone can (and does) send email to any user, only people who are on a user's buddy-list may send this user instant messages. Thus, since messages are received from a select group of known contacts, one can assume that these contacts are bound by a social contract that will deter them from abusing their buddy. Of course, an insensitive buddy who abuses QnA can ultimately be blocked (through an invisible list), or removed them from the buddy-list altogether. One could argue that, if used sensitively, allowing users to increase the salience of their messages occasionally may be an additional benefit of QnA.

7.5.3 Notifications vs. Message Previews

Following a few inquiries as to the design of QnA I have added the option to allow QnA to present users with a preview of the message rather than notifications of the arrival of a question or an answer (see Figure 7.1d). It is my opinion, however, that one's ability to ignore a question once its content is known is greatly reduced compared to one's ability to ignore a question based solely on the identity of its sender (and the relevance of this sender to one's ongoing task). Indeed, to the best of my knowledge, the option for viewing a preview of the message is used by none of the users of QnA.

7.5.4 QnA in Multi-Monitor Conditions

One interesting an unanticipated benefit of QnA was described by a user who often used more than one monitor simultaneously. This user stated that since he used IM mostly on his secondary monitor, QnA notifications helped him attend to new and ongoing IM conversations. A related field study of users of multiple monitors found that users will often keep communication applications in the peripheral monitor (such as IM or email), while keeping their main tasks in the primary monitor (Grudin, 2001).

7.6 Summary & Future Work

In this chapter, I have presented QnA, a tool that augments a commercial IM client to allow users to maintain a flow of work by providing salient notifications of incoming messages that may deserve their attention. In particular, QnA focuses on incoming questions and answers as those messages are typically associated with a buddy waiting for a response (in the case of questions), or messages the user is waiting for (in the case of answers). Preliminary results from a quantitative evaluation suggest that QnA can indeed affect users' interaction with IM, allowing them to read messages faster when those messages contain questions. Confirming these findings with a more extensive quantitative evaluation is in need. Furthermore, a qualitative evaluation is needed for examining the effect of QnA on users' attitudes to IM.

While the set of rules used for determining if a message contains a question (and rules for messages that are not questions) is continuously expanded and refined, future versions of QnA may also include the option to allow users to create custom rules. Finally, I am interested in allowing users to add buddies to an "ignore" list that prevents QnA from displaying notifications for those buddies.

CHAPTER EIGHT

Conclusions

At the heart of the work presented in this dissertation is the notion that interpersonal communication is a good, necessary, and desirable element of our lives. This work recognizes, as many have before, than when communication is mediated by technology, the interaction between the communication and the context into which it arrives can make communication a burden.

In order to increase our understanding of the use of communication tools, and in order to enable the creation of enhanced communication technology, I have detailed in this dissertation a collection of quantitative explorations of aspects of technology-mediated communication use, described the creation of a number of statistical predictive models, and developed and provided an initial evaluation of a communication enhancement tool. Together, these elements present a rich, interdisciplinary investigation of technology-mediated semi-synchronous communication, with contributions in both theoretical and applied domains. The following sections provide a review of many of the central findings within each of these categories:

8.1 Applied Contributions

In this dissertation I have described the development of tools and models necessary for the creation of enhanced communication systems. I've described the creation of models that are able to accurately predict, based on activity and past interaction, a user's responsiveness to incoming IM communication. I've presented models that are able to predict, based on past communication patterns, the relationship between communication partners. And I have described a tool that allows users to easily identify messages that may require their quick response, by that helping them balance their responsiveness and their performance on their ongoing tasks.

In Chapter 3, I described the process of creating a set of statistical models that are able to predict a user's responsiveness to incoming instant messages. More specifically, I described models that predict, with very high accuracy, responsiveness to attempts to initiate new communication (arguably, the point in a conversation for which predictions of responsiveness are most useful). These models were based on field-data collected over month-long periods in participants' natural surroundings. Such models could be used to automatically provide different "traditional" online-status indicators to different buddies. Alternatively, models can be used to increase the salience of incoming messages that may deserve immediate attention if responsiveness is predicted to be low. Models could also be used by a system that will show a list of potentially responsive buddies to users who are looking for help or support, while hiding others.

In Chapter 4, I also described the creation of models that predict responsiveness to incoming IM without using information about the buddy. I showed that these models (referred to as *Buddy-Independent*) were able to predict responsiveness with accuracy that is significantly higher than the prior probability, and with only slightly (and not significantly) lower accuracy than the first set of models. Buddy-independent models are of particular interest from a practical standpoint. Models that use the full feature-set (knowing, for example, how much time has passed since the last time a message was exchanged with a specific buddy) may predict, at the same time, different levels of responsiveness to different buddies. In contrast, buddy-independent models are oblivious to information about the source of the message, and will predict, at any point in time, the same level of responsiveness to all buddies, basing the prediction only on information that is "local" to the user. In the design of a system that uses models of responsiveness, the system designer will need to carefully consider whether to provide a unified prediction of responsiveness to all buddies (using *buddy-independent* models) or whether additional benefit may be gained by providing different predictions to different buddies

An examination of the interaction between the time that has passed since the arrival of a message and the likelihood of a response was presented in Chapter 4. Unlike the models presented in Chapter 3, which aim to provide benefit through predictions of responsiveness prior to the delivery of a message, in this chapter I examined forecasts of responsiveness to messages that have already been sent and while the sender is waiting for a response. This investigation of response-likelihood may provide benefit beyond merely adding to the body of research on the probability distribution of asynchronous communication; rather it may provide multiple different and potentially applicable views into the underlying distribution of IM responsiveness.

In Chapter 6, I presented statistical models that classify the relationship between conversation partners based on past communication. One of the models described was able to classify, with 79.3% accuracy, whether a user and a buddy are in a work or social relationship. This accuracy is impressively high considering that only basic characteristics of communication were used for the classification, without knowledge of the actual content of messages.

Finally, in Chapter 7, I presented a tool that allows users to balance their performance on ongoing tasks with their responsiveness to incoming messages. Specifically, this tool helps users identify messages that require quick responses as well as those that they are waiting for from others. The preliminary evaluation, presented at the end of this chapter, suggested this tool's effectiveness in influencing responsiveness to messages that require it.

8.2 Theoretical Contributions

One of the primary goals of the work described in this dissertation is to advance our understanding of interpersonal communication as it is mediated by technology, particularly for semi-synchronous communication. This understanding allows us to understand people as they engage in communication, but also allows us to guide our efforts in designing novel communication tools.

In Chapter 5, I described results from an in-depth analysis of factors that affect responsiveness to incoming instant messages. Through this analysis I was able to advance our understanding of responsiveness and its relationship with a user's availability. While this work describes investigation of responsiveness in a single medium (IM), the general classes of measures that were investigated – context, communication, and content – are not at all unique to IM, but generalize to other forms of interpersonal communication. An investigation of responsiveness as it is manifested in other media (and as different media interact), would be interesting and beneficial.

In Chapter 6, I described an analysis of the effect of the relationship between IM communication partners on basic features of their IM communication. I presented, for example, a number of results that suggest that, while IM sessions with social contacts are longer in duration, users focus less of their undivided attention, on average, to these sessions. This work on IM and interpersonal relationships extends previous research that showed the effect of interpersonal relationships on face-to-face and phone communication. This work also complements previous research that described the effect of frequency of communication on basic characteristics of communication in both synchronous and asynchronous mediums.

8.3 Discussion

To conclude, I wish to discuss, briefly, a number of issues that have come up during the course of this thesis work.

8.3.1 Responsiveness, Norms, and Culture

The link between *availability* and *responsiveness*, presented in this dissertation, is likely to be influenced by cultural and normative elements. (Consequently, when discussing this link, I have offered the term *demonstrated availability* to describe availability as it is enacted, rather than the way it is desired.) Indeed, cultural differences have been shown by prior research to result in differences in communication and the use of communication technology (for example, Massey, Hung, Montoya-Weiss, & Ramesh, 2001; Setlock, Fussell, & Neuwirth, 2004; Choi, Lee, Kim, & Jeon, 2005; Kayan, Fussell, & Setlock, 2006). Organizational norms, too, can have great impact on the use and adoption of communication technology (Kraut, Rice, Cool, & Fish, 1998). In the realm of IM communication, for example, organizational norms and choices may affect basic critical elements of the medium and have great impact on its culture of use. An organization (such as IBM, for example) may require an employee's electronic identity, associated with their email address and IM name, to be visible to all other employees. By that, the organization is mandating that an employee is accessible through IM to any other member of the organization, not only to this employee's close network. This, in turn, means that an incoming message can no longer be approximated to originate from a small set of contacts. While messages from unfamiliar contacts are, in practice, few and far apart, they do occur, resulting in a change in attitude towards IM, with some users electing to avoid it altogether.

Understanding the culture and normative settings is thus critical when introducing predictive models into communication technology. (Keep in mind that the models or responsiveness presented in Chapter 3 are indifferent to a user's cultural and normative settings – such models learn to predict the act of responsiveness from the user's observed actions, whether or not these actions are influenced by culture and norms.) A system that uses predictive models to provide enhanced contextual awareness, for example, may be welcome in one culture or organization, but may result in users avoiding such a communication tool in another. It is thus necessary to examine the impact of culture and norms on both demonstrated and desired availability in order to better understand the potential impact of predictive models in different settings.

8.3.2 Evaluating Inaction

The work presented in this dissertation aims to assist people in finding opportune moments for successful communication and reducing disruptive communication. Put differently, this work aims to discourage, remove, or assist in avoiding an interruption, that is, work that has the desired result of *inaction* rather than *action.*

As such, this research joins a wide range of research work looking at preventing, removing, or discouraging behavior. These works include other tools aimed at reducing interruptions through indicators of availability, work aimed at lowering users' energy consumption, work aimed at changing eating and exercise habits, etc.

The difficulty of evaluating this type of work is worth discussing. While in a laboratory experiment, a person's goals and intentions can be controlled and manipulated, evaluating tools and methods, aimed at creating *inaction* rather than *action*, in the field, is difficult since the researcher does not have access to the participant's intent. That is, observed action does not necessarily mean that an intervention is failing, while not observing an action cannot immediately be attributed to successful intervention – one has no realistic way of knowing that an action was intended in the first place but discouraged by the tool.

To illustrate the difficulty of evaluating inaction, consider the following example:

A researcher is interested in evaluating the effectiveness of a posted sign that asks passers-by not to litter. Such evaluation would be impossible to do through mere observation. Since the researcher does not know the intentions of a person walking past the sign, they cannot conclude that the person *avoided* littering due to the sign, because the researcher doesn't know that the person intended to litter in the first place. Furthermore, a person who did intend to litter but did not, might have done so for reasons other than the posted sign. On the other hand, observing a person littering does not immediately implicate that posted signs are ineffective in general (although the particular sign may be suspect). Indeed, it is possible that the person did not see the sign or understand its meaning.

It would thus seem that evaluations of tools and research, whose goal is user inaction, should combine observations, probes of user's intentions, and qualitative measures of change over time. The creation of a framework for the evaluation of inaction would be both interesting, as well as a very useful research effort.

8.4 Closing Remarks

In conclusion, communication technology is maturing and with it, its users. The young adults who have been using IM and mobile telephony for their social communication for over a decade are now joining the workforce. Thus, better communication tools and a better understanding of the factors that influence the use of these tools are needed.

In this dissertation I argued for a research approach that combines the creation of communication tools with investigation of the factors these tools aim to address. It is through such a combined approach that we may understand the successes and failures of our tools, on the one hand, and be confident that our tools solve the right problems, on the other. The work presented in this dissertation is an important step towards reaching these goals.

References

Adamczyk, P. D., & Bailey, B. P. (2004). If not now, when?: the effects of interruption at different moments within task execution. In Proceedings of *the ACM Conference on Human Factors in Computing Systems (CHI 2004)*, pp. 271-278. ACM Press.

Adler, A., & Henderson, A. (1994). A Room of Our Own: Experiences from a Direct Office Share. In Proceedings of *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 1994) (CHI 1994)*, pp. 138-144. ACM Press.

Agichtein, E., S., L., & Gravano, L. (2001). Learning search engine specific query transformations for question answering. In Proceedings of *the tenth international conference on World Wide Web (WWW 2001)*, pp. 169-178. ACM Press.

Avrahami, D., Fogarty, J., & Hudson, S. E. (2007a). Biases in Human Estimation of Interruptibility: Effects and Implications for Practice. In Proceedings of *the ACM Conference on Computer Supported Cooperative Work (CHI 2007)*, pp. 50-60. ACM Press.

Avrahami, D., Gergle, D., Hudson, S. E., & Kiesler, S. (2007b). Improving the match between callers and receivers: A study on the effect of contextual information on cell phone interruptions. *Behaviour & Information Technology, 26*(3).

Avrahami, D., & Hudson, S. E. (2004). QnA: Augmenting an Instant Messaging Client to Balance User Responsiveness and Performance. In Proceedings of *the ACM Conference on Computer Supported Cooperative Work (CSCW 2004)*, pp. 515-518. ACM Press.

Avrahami, D., & Hudson, S. E. (2006a). Communication Characteristics of Instant Messaging: Effects and Predictions of Interpersonal Relationship. In Proceedings of *the ACM Conference on Computer Supported Cooperative Work (CSCW 2006)*, pp. 505-514. ACM Press.

Avrahami, D., & Hudson, S. E. (2006b). Responsiveness in Instant Messaging: Predictive Models Supporting Inter-Personal Communication. In Proceedings of *the ACM Conference on Human Factors in Computing Systems (CHI 2006)*, pp. 731-740. ACM Press.

Bailey, B. P., Adamczyk, P. D., Chang, T. Y., & Chilson, N. A. (2006). A Framework for Specifying and Monitoring User Tasks. *Journal of Computers in Human Behavior, 22*(4), 685-708.

Bailey, B. P., Konstan, J. A., & Carlis, J. V. (2000). Measuring the effects of interruptions on task performance in the user interface. In Proceedings of *IEEE International Conference on Systems, Man, and Cybernetics (SMC 2000)*, pp. 757-762. IEEE.

Bailey, B. P., Konstan, J. A., & Carlis, J. V. (2001). The effects of interruptions on task performance, annoyance, and anxiety in the user interface. In Proceedings of *Human-Computer Interaction (INTERACT 2001)*, pp. 593-601. Amsterdam: IOS Press.

Bao, L., & Intille, S. S. (2004). Activity Recognition from User-Annotated Acceleration Data. In Proceedings of *the International Conference on Pervasive Computing (PERVASIVE 2004)*, pp. 1-17.

Barabási, A. L. (2005). The origin of bursts and heavy tails in human dynamics. *Nature, 435*(7039), 207-211.

Bartram, L., Ware, C., & Calvert, T. (2003). Moticons: Detection, distraction and task. *International Journal of Human-Computer Studies, 58*(5), 515-545.

Begole, J. B., Matsakis, N. E., & Tang, J. C. (2004). Lilsys: Sensing Unavailability. In Proceedings of *the ACM Conference on Computer Supported Cooperative Work (CSCW 2004)*, pp. 511-514.

Begole, J. B., Tang, J. C., & Hill, R. (2003). Rhythm Modeling, Visualizations, and Applications. In Proceedings of *the ACM Symposium on User Interface Software and Technology (UIST 2003)*, pp. 11-20. ACM Press.

Begole, J. B., Tang, J. C., Smith, R. B., & Yankelovich, N. (2002). Work Rhythms: Analyzing Visualizations of Awareness Histories of Distributed Groups. In Proceedings of *the ACM Conference on Computer Supported Cooperative Work (CSCW 2002)*, pp. 334-343. ACM Press.

Bellotti, V., & Edwards, K. (2001). Intelligibility and Accountability: Human Considerations in Context-Aware Systems. *Human-Computer Interaction, 16*, 193-212.

Bly, S. A., Harrison, S. R., & Irwin, S. (1993). Media Spaces: Bringing People Together in a Video, Audio, and Computing Environment. *Communications of the ACM, 36*(1), 28-46.

Bradner, E., Kellogg, W. A., & Erickson, T. (1999). The adoption and use of 'Babble': A field study of chat in the workplace. In Proceedings of *the European conference on Computer Supported Cooperative Work (ECSCW 1999)*, pp. 139-158. Springer Publishing.

Burke, M., Joyce, E., Kim, T., Anand, V., & Kraut, R. E. (2007). Introductions and Requests: Rhetorical Strategies That Elicit Response in Online Communities. In Proceedings of *Third International Conference on Communities & Technologies 2007 (C&T 2007)*. Springer.

Cerulean Studios - Trillian Pro. 2006, from http://www.trillian.cc

Choi, B., Lee, I., Kim, J., & Jeon, Y. (2005, April 02 - 07, 2005). A qualitative cross-national study of cultural influences on mobile data service design. . In Proceedings of *the SIGCHI Conference on Human Factors in Computing Systems (CHI 2005)*, pp. 661-670. ACM Press.

Clark, H. H. (1992). *Arenas of language use*. Chicago, IL Stanford, CA: The University of Chicago Press Center for the Study of Language & Information.

Crown, C. L. (1982). Impression Formation and the Chronography of Dyadic Interactions. In M. Davis (Ed.), *Interaction Rhythms* (pp. 225-248). New York, NY: Human Sciences Press Inc.

Csikszentmihalyi, M., Larson, R., & Prescott, S. (1977). The ecology of adolescent activity and experience. *Journal of Youth and Adolescence, 6*(3), 281-294.

Cummings, J., & Ghosh, T. (under review). Dealing with distance in team networks: A relational model of communication technology use.

Cutrell, E., Czerwinski, M., & Horvitz, E. (2001). Notification, Disruption, and Memory: Effects of Messaging Interruptions on Memory and Performance. In Proceedings of *Human-Computer Interaction (INTERACT 2001)*, pp. 263-269. IOS Press.

Czerwinski, M., Cutrell, E., & Horvitz, E. (2000a). Instant Messaging and Interruptions: Influence of Task Type on Performance. In Proceedings of *the Australian Conference on Computer-Human Interaction (OZCHI 2000)*, pp. 356-361.

Czerwinski, M., Cutrell, E., & Horvitz, E. (2000b). Instant Messaging: Effects of Relevance and Time. In Proceedings of *the British HCI Group Annual Conference (HCI 2000)*, pp. 71-76.

Czerwinski, M., Horvitz, E., & Wilhite, S. (2004). A Diary Study of Task Switching and Interruptions. In Proceedings of *the ACM Conference on Human Factors in Computing Systems (CHI 2004)*, pp. 175-182.

Dabbish, L. (2007). *Coordinating Initiation and Response in Computer-Mediated Communication.* Unpublished Ph.D. Dissertation, Carnegie Mellon University.

Dabbish, L., & Kraut, R. E. (2004). Controlling Interruptions: Awareness Displays and Social Motivation for Coordination. In Proceedings of *the ACM Conference on Computer Supported Cooperative Work (CSCW 2004)*, pp. 182-191.

Dabbish, L., Kraut, R. E., Fussell, S. R., & Kiesler, S. (2005). Understanding email use: Predicting action on a message. In Proceedings of *the ACM Conference on Human Factors in Computing Systems (CHI 2005)*, pp. 691-700. ACM Press.

Dourish, P., & Bly, S. (1992). Portholes: Supporting Awareness in a Distributed Work Group. In Proceedings of *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 1992) (CHI 1992)*, pp. 541-547. ACM Press.

Duck, S., Turr, D., Hurst, M., & Strejc, H. (1991). Some evident truths about conversations in everyday relationships: All communications are not created equal. *Human Communication Research, 18*(2), 228-267.

Eyrolle, H., & Cellier, J. M. (2000). The effects of interruptions in work activity: Field and laboratory results. *Applied Ergonomics, 31*(5), 537-543.

Feldstein, S. (1982). Impression Formation in Dyads: The Temporal Dimension. In M. Davis (Ed.), *Interaction Rhythms* (pp. 207-224). New York, NY: Human Sciences Press Inc.

Fish, R. S., Kraut, R. E., Root, R. W., & Rice, R. E. (1992). Video as a Technology for Informal Communication. In Proceedings of *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 1992)*, pp. 37-48. ACM Press.

Fogarty, J., Au, C., & Hudson, S. E. (2006). Sensing from the Basement: A Feasibility Study of Unobtrusive and Low-Cost Home Activity Recognition. In Proceedings of *the ACM Symposium on User Interface Software and Technology (UIST 2006)*, pp. 91-100. ACM Press.

Fogarty, J., & Hudson, S. E. (2007). Toolkit support for developing and deploying sensor-based statistical models of human situations. In Proceedings of *the SIGCHI Conference on Human Factors in Computing Systems (CHI 2007)*, pp. 135-144. ACM Press.

Fogarty, J., Hudson, S. E., Atkeson, C. G., Avrahami, D., Forlizzi, J., Kiesler, S., Lee, J. C., & Yang, J. (2005a). Predicting Human Interruptibility with Sensors. *ACM Transactions on Computer-Human Interaction (TOCHI), 12*(1), 119-146.

Fogarty, J., Hudson, S. E., & Lai, J. (2004a). Examining the Robustness of Sensor-Based Statistical Models of Human Interruptibility. In Proceedings of *the ACM Conference on Human Factors in Computing Systems (CHI 2004)*, pp. 207-214.

Fogarty, J., Ko, A. J., Aung, H. H., Golden, E., Tang, K. P., & Hudson, S. E. (2005b). Examining Task Engagement in Sensor-Based Statistical Models of Human Interruptibility. In Proceedings of *the ACM Conference on Human Factors in Computing Systems (CHI 2005)*, pp. 331-340. ACM Press.

Fogarty, J., Lai, J., & Christensen, J. (2004b). Presence versus availability: The design and evaluation of a context-aware communication client. *IJHCS, 61*(3).

Freund, Y., & Schapire, R. E. (1996). Experiments with a new boosting algorithm. In Proceedings of *the Thirteenth International Conference on Machine Learning (ICML'96)*, pp. 148-156. Morgan Kaufmann.

Gergle, D., Millen, D., Kraut, R. E., & Fussell, S. R. (2004). Persistence Matters: Making the Most of Chat in Tightly-Coupled Work. In Proceedings of *the ACM Conference on Human Factors in Computing Systems (CHI 2004)*, pp. 431-438. ACM Press.

Ghosh, T., Yates, J., & Orlikowski, W. J. (2004). Using communication norms for coordination: Evidence from a distributed team. In Proceedings of *the International Conference on Information Systems (ICIS 2004)*, pp. 115-128.

Gillie, T., & Broadbent, D. (1989). What Makes Interruptions Disruptive? A Study of Length, Similarity, and Complexity. *Psychological Research, 50*, 243-250.

Gluck, J., Bunt, A., & McGrenere, J. (2007). Matching Attentional Draw with Utility in Interruption. In Proceedings of *the ACM Conference on Human Factors in Computing Systems (CHI 2007)*. ACM Press.

Goldsmith, D. J., & Baxter, L. A. (1996). Constituting Relationships in Talk: A Taxonomy of Speech Events in Social and Personal Relationships. *Human Communication Research, 23*(1), 87-114.

Gonzalez, V., & Mark, G. (2004, April 26-29). "Constant, Constant, Multi-tasking Craziness": Managing Multiple Working Spheres. In Proceedings of *the ACM Conference on Human Factors in Computing Systems (CHI 2004)*.

Grinter, R., & Palen, L. (2002). Instant Messaging in Teen Live. In Proceedings of *the ACM Conference on Computer Supported Cooperative Work (CSCW 2002)*, pp. 21-30. ACM Press.

Grudin, J. (2001). Partitioning Digital Worlds: Focal and Peripheral Awareness in Multiple Monitor Use. In Proceedings of *the ACM Conference on Human Factors in Computing Systems (CHI 2001)*, pp. 458-465. ACM Press.

Hafner, K. (2003, August 14). All Quiet on Campus Save the Click of Keys. *New York Times*.

Handel, M., & Herbsleb, J. D. (2002). What Is Chat Doing in the Workplace? In Proceedings of *the ACM Conference on Computer Supported Cooperative Work (CSCW 2002)*, pp. 1-10. ACM Press.

Herbsleb, J. D., Atkins, D. L., Boyer, D. G., Handel, M., & Finholt, T. A. (2002). Introducing Instant Messaging and Chat in the Workplace. In Proceedings of *the ACM Conference on Human Factors in Computing Systems (CHI 2002)*, pp. 171-178.

Ho, J., & Intille, S. S. (2005). Using Context-Aware Computing to Reduce the Perceived Burden of Interruptions from Mobile Devices. In Proceedings of *the ACM Conference on Human Factors in Computing Systems (CHI 2005)*, pp. 909-918.

Horvitz, E., & Apacible, J. (2003). Learning and Reasoning about Interruption. In Proceedings of *the International Conference on Multimodal Interfaces (ICMI 2003)*, pp. 20-27.

Horvitz, E., Apacible, J., & Subramani, M. (2005a). Balancing Awareness and Interruption: Investigation of Notification Deferral Policies. In Proceedings of *the Tenth Conference on User Modeling (UM 2005)*.

Horvitz, E., Jacobs, A., & Hovel, D. (1999). Attention-Sensitive Alerting. In Proceedings of *Conference on Uncertainty and Artificial Intelligence (UAI 1999)*, pp. 305-313.

Horvitz, E., Kadie, C. M., Peak, D., & Hovel, D. (2003). Models of Attention in Computing and Communications: From Principles to Applications. *Communications of the ACM, 46*(3), 52-59.

Horvitz, E., Koch, P., & Apacible, J. (2004). BusyBody: Creating and Fielding Personalized Models of the Cost of Interruption. In Proceedings of *the ACM Conference on Computer Supported Cooperative Work (CSCW 2004)*, pp. 507-510.

Horvitz, E., Koch, P., Kadie, C. M., & Jacobs, A. (2002). Coordinate: Probabilistic Forecasting of Presence and Availability. In Proceedings of *the Conference on Uncertainty in Artificial Intelligence (UAI 2002)*, pp. 224-233.

Horvitz, E., Koch, P., Sarin, R., Apacible, J., & Subramani, M. (2005b, 2005-09-20). Bayesphone: Precomputation of Context-Sensitive Policies for Inquiry and Action in Mobile Devices. In Proceedings of *User Modeling (UM 2005)*, pp. 251-260. Springer.

Hudson, J. M., Christensen, J., Kellogg, W. A., & Erickson, T. (2002). "I'd be overwhelmed, but it's just one more thing to do": Availability and Interruption in Research Management. In Proceedings of *the ACM Conference on Human Factors in Computing Systems (CHI 2002)*, pp. 97-104. ACM Press.

Hudson, S. E., Fogarty, J., Atkeson, C. G., Avrahami, D., Forlizzi, J., Kiesler, S., Lee, J. C., & Yang, J. (2003). Predicting Human Interruptibility with Sensors: A Wizard of Oz Feasibility Study. In Proceedings of *the ACM Conference on Human Factors in Computing Systems (CHI 2003)*, pp. 257-264. ACM Press.

Hudson, S. E., & Smith, I. (1996). Techniques for Addressing Fundamental Privacy and Disruption Tradeoffs in Awareness Support Systems. In Proceedings of *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW 1996)*, pp. 248-257. ACM Press.

Iqbal, S. T., Adamczyk, P. D., Zheng, X. S., & Bailey, B. P. (2005). Towards an Index of Opportunity: Understanding Changes in Mental Workload During Task Execution. In Proceedings of *the ACM Conference on Human Factors in Computing Systems (CHI 2005)*, pp. 311-320.

Iqbal, S. T., & Bailey, B. P. (2006). Leveraging characteristics of task structure to predict the cost of interruption. In Proceedings of *the ACM Conference on Human Factors in Computing Systems (CHI 2006)*, pp. 741-750. ACM Press.

Iqbal, S. T., & Bailey, B. P. (2007). Understanding and Developing Models for Detecting and Differentiating Breakpoints during Interactive Tasks. In Proceedings of *the ACM Conference on Human Factors in Computing Systems (CHI 2007)*. ACM Press.

Iqbal, S. T., & Horvitz, E. (2007a). Conversations Amidst Computing: A Study of Interruptions and Recovery of Task Activity. In Proceedings of *the Eleventh Conference on User Modeling (UM 2007)*.

Iqbal, S. T., & Horvitz, E. (2007b). Disruption and Recovery of Computing Tasks: Field Study, Analysis and Directions. In Proceedings of *the ACM Conference on Human Factors in Computing Systems (CHI 2007)*. ACM Press.

Isaacs, E., Walendowski, A., Whittaker, S., Schiano, D. J., & Kamm, C. (2002). The Character, Functions, and Styles of Instant Messaging in the Workplace. In Proceedings of *the ACM Conference on Computer Supported Cooperative Work (CSCW 2002)*, pp. 11-20. ACM Press.

Jones, E. E., & Nisbett, R. E. (1971). The actor and observer: Divergent perceptions of the causes of behavior. In E. E. Jones, D. E. Kanouse, H. H. Kelly, R. E. Nisbett, S. Valins & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior*.

Jung, Y., Persson, P., & Blom, J. (2005, April 02 - 07, 2005). DeDe: design and evaluation of a context-enhanced mobile messaging system. In Proceedings of *the SIGCHI Conference on Human Factors in Computing Systems (CHI 2005)*, pp. 351-360. ACM Press.

Kalman, Y. M., & Rafaeli, S. (2005). Email chronemics: unobtrusive profiling of response times. In Proceedings of *the 38th Annual Hawaii International Conference on System Sciences (HICSS 2005)*, pp. 108b.

Kalman, Y. M., Ravid, G., Raban, D. R., & Rafaeli, S. (2006a). Pauses and response latencies: A chronemic analysis of asynchronous CMC. *Journal of Computer-Mediated Communication, 12*(1).

Kalman, Y. M., Ravid, G., Raban, D. R., & Rafaeli, S. (2006b). Speak *now* or forever hold your peace: power law chronemics of turn-taking and response in asynchronous CMC. In Proceedings of *Annual Conference of the International Communication Association (ICA 2006)*.

Kapoor, A., & Horvitz, E. (2007). Principles of Lifelong Learning for Predictive User Modeling. In Proceedings of *the Eleventh Conference on User Modeling (UM 2007)*.

Kapoor, A., Horvitz, E., & Basu, S. (2007). Selective Supervision: Guiding Supervised Learning with Decision-Theoretic Active Learning. In Proceedings of *International Joint Conference on Artificial Intelligence (IJCAI 2007)*, pp. 877-882.

Kayan, S., Fussell, S. R., & Setlock, L. D. (2006). Cultural differences in the use of instant messaging in Asia and North America In Proceedings of *the ACM Conference on Computer Supported Cooperative Work (CSCW 2006)*, pp. 525-528 ACM Press.

Kern, N., Antifakos, S., Schiele, B., & Schwaninger, A. (2004). A model for human interruptability: Experimental evaluation and automatic estimation from wearable sensors. In Proceedings of *the Eighth International Symposium on Wearable Computers (ISWC 2004)*, pp. 158-165. IEEE Computer Society.

Khalil, A., & Connelly, K. (2006, November 04 - 08, 2006). Context-aware telephony: privacy preferences and sharing patterns. In Proceedings of *the 20th Anniversary Conference on Computer Supported Cooperative Work (CSCW '06)*, pp. 469-478. ACM Press.

Kohavi, R., & John, G. H. (1997). Wrappers for Feature Subset Selection. *Artificial Intelligence, 97*(1-2), 273-324.

Kraut, R. E., & Attewell, P. (1997). Media use and organizational knowledge: Electronic mail in a global corporation. In S. Kiesler (Ed.), *Research milestones on the information highway*. Mahwah, NJ: Erlbaum.

Kraut, R. E., Fish, R. S., Root, R. W., & Chalfonte, B. L. (1990). Informal Communication in Organizations: Form, Function, and Technology. In S. Oskamp & S. Spacapan (Eds.), *People's Reactions to Technology* (pp. 145-199). Newberry Park: Sage.

Kraut, R. E., Rice, R. E., Cool, C., & Fish, R. S. (1998). Varieties of social influence: The role of utility and norms in the success of a communication medium. *Organizational Science, 9*(4), 437-453.

Leskovec, J., & Horvitz, E. (2007). Worldwide Buzz: Planetary-Scale Views on an Instant-Messaging Network *Microsoft Research Technical Report, MSR-TR-2006-186*.

Mahowald, R. P. (2005). *Worldwide Enterprise Instant Messaging Applications 2005-2009 Forecast and 2004 Vendor Shares: Clearing the Decks for Substantial Growth*: IDC Market Analysis.

Mark, G., Gonzalez, V., & Harris, J. (2005). No Task Left Behind? Examining the Nature of Fragmented Work. In Proceedings of *the ACM Conference on Human Factors in Computing Systems (CHI 2005)*, pp. 321-330. ACM Press.

Massey, A. P., Hung, Y. C., Montoya-Weiss, M., & Ramesh, V. (2001, September 30, 2001). When culture and style aren't about clothes: perceptions of task-technology "fit" in global virtual teams. In Proceedings of *the international ACM SIGGROUP Conference on Supporting Group Work (GROUP 2001)*, pp. 207-213. ACM Press.

McCrickard, D. S., Catrambone, R., Chewar, C. M., & Stasko, J. T. (2003). Establishing trade-offs that leverage attention for utility: Empirically evaluating information display in notification systems. *International Journal of Human-Computer Studies, 58*(5), 547-582.

McCrickard, D. S., & Chewar, C. M. (2003). Attuning notification design to user goals and attention costs. *Communications of ACM, 46*(3), pp. 67-72.

McFarlane, D. C. (2002). Comparison of Four Primary Methods for Coordinating the Interruption of People in Human-Computer Interaction. *Human-Computer Interaction, 17*(1), 63-139.

McFarlane, D. C., & Latorella, K. A. (2002). The Scope and Importance of Human Interruption in Human-Computer Interaction Design. *Human-Computer Interaction, 17*(1), 1-61.

McLaughlin, M. L., & Cody, M. J. (1982). Awkward silences: Behavioral antecedents and consequences of the conversational lapse. *Human Communication Research, 8*(4), 299–316.

Milewski, A. E., & Smith, T. M. (2000). Providing Presence Cues to Telephone Users. In Proceedings of *the ACM Conference on Computer Supported Cooperative Work (CSCW 2000)*, pp. 89-96. ACM Press.

Mirabilis Inc. *ICQ*, 2007, from www.mirabilis.com

Miyata, Y., & Norman, D. A. (1986). Psychological issues in support of multiple activities. In D. A. Norman & S. W. Draper (Eds.), *User Centered Systems Design: New Perspectives on Human-Computer Interaction* (pp. 265-284). Hillsdale: Lawrence Erlbaum Associates.

Monk, C. A. (2004). The effect of frequent versus infrequent interruptions on primary task resumption. In Proceedings of *the 48th Annual Meeting of the Human Factors and Ergonomics Society (HFES 2004)*, pp. 295-299. Santa Monica: Human Factors and Ergonomics Society.

Monk, C. A., Boehm-Davis, D. A., & Trafton, J. G. (2002). The attentional costs of interrupting task performance at various stages. In Proceedings of *the 46th Annual Meeting of the Human Factors and Ergonomics Society (HFES 2002)*, pp. 1824-1828. Human Factors and Ergonomics Society.

Mynatt, E., & Tullio, J. (2001). Inferring Calendar Event Attendance. In Proceedings of *the International Conference on Intelligent User Interfaces (IUI 2001)*, pp. 121-128. ACM Press.

Nagel, K. S., Hudson, J. M., & Abowd, G. (2004). Predictors of Availability in Home Life Context-Mediated Communication. In Proceedings of *the ACM Conference on Computer Supported Cooperative Work (CSCW 2004)*, pp. 497-506.

Nardi, B. A., Whittaker, S., & Bradner, E. (2000). Interaction and Outeraction: Instant Messaging in Action. In Proceedings of *the ACM Conference on Computer Supported Cooperative Work (CSCW 2000)*, pp. 79-88. ACM Press.

O'Conaill, B., & Frohlich, D. (1995). Timespace in the Workplace: Dealing with Interruptions. In Extended Abstracts of *the ACM Conference on Human Factors in Computing Systems (CHI 1995)*, pp. 262-263. ACM Press.

Pedersen, E. R. (2001). Calls.calm: Enabling Caller and Callee to Collaborate. In Extended Abstracts of *the ACM Conference on Human Factors in Computing Systems (CHI 2001)*, pp. 235-236. ACM Press.

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Linguistic inquiry and word count (LIWC)*. Mahwah, NJ: Erlbaum Publishers.

Perlow, L. A. (1999). The Time Famine: Toward a Sociology of Work Time. *Administrative Science Quarterly, 44*(1), 57-81.

Rennecker, J., & Godwin, L. (2005). Delays and interruptions: A self-perpetuating paradox of communication technology use. *Information and Organization, 15*(3), 247-266.

Robertson, T. J., Prabhakararao, S., Burnett, M., Cook, C., Ruthruff, J. R., Beckwith, L., & Phalgune, A. (2004). Impact of Interruption Style on End-User Debugging. In Proceedings of *the ACM Conference on Human Factors in Computing Systems (CHI 2004)*, pp. 287-294. ACM Press.

Schegloff, E. A., & Sacks, H. (1973). Opening and Closing. *Semiotica, 8*(4), 289-327.

Schmidt, A., Takaluoma, A., & Mäntyjärvi, J. (2000). Context-Aware Telephony Over WAP. *Personal and Ubiquitous Computing, 4*(4), 225-229.

Scollon, C. N., Kim-Prieto, C., & Diener, E. (2003). Experience Sampling: Promises and Pitfalls, Strengths and Weaknesses. *Journal of Happiness Studies, 4*(1), 5-34.

Setlock, L. D., Fussell, S. R., & Neuwirth, C. (2004, November 06 - 10, 2004). Taking it out of context: collaborating within and across cultures in face-to-face settings and via instant messaging. In Proceedings of *the ACM Conference on Computer Supported Cooperative Work (CSCW 2004)*, pp. 604-613. ACM Press.

Slatalla, M. (1999, Sept. 23). The office meeting that never ends. *New York Times*.

Tang, J. C., Yankelovich, N., Begole, J., Van Kleek, M., Li, F., & Bhalodia, J. (2001). ConNexus to Awarenex: Extending Awareness to Mobile Users. In Proceedings of *the ACM Conference on Human Factors in Computing Systems (CHI 2001)*, pp. 221-228. ACM Press.

Tullio, J., Dey, A. K., Chalecki, J., & Fogarty, J. (2007). How it works: a field study of non-technical users interacting with an intelligent system. In Proceedings of *the SIGCHI Conference on Human Factors in Computing Systems (CHI 2007)*, pp. 31-40. ACM Press.

Tullio, J., Goecks, J., Mynatt, E. D., & Nguyen, D. H. (2002, October 27 - 30, 2002). Augmenting shared personal calendars. In Proceedings of *the 15th Annual ACM Symposium on User interface Software and Technology (UIST 2002)*, pp. 11-20. ACM Press.

Tyler, J. R., & Tang, J. C. (2003). When Can I Expect an Email Response? A Study of Rhythms in Email Usage. In Proceedings of *the European conference on Computer Supported Cooperative Work (ECSCW 2003)*, pp. 238-258. Oulu Univ. Press.

Voida, A., Newstetter, W. C., & Mynatt, E. D. (2002). When Conventions Collide: The Tensions of Instant Messaging Attributed. In Proceedings of *the ACM Conference on Human Factors in Computing Systems (CHI 2002)*, pp. 187-194. ACM Press.

Wellman, B. (2001). Physical Place and Cyber Place: The Rise of Personalized Networking. *International Journal of Urban and Regional Research, 25*(2), 227-252.

Whittaker, S. J., Frohlich, D., & Daly-Jones, O. (1994). Informal Workplace Communication: What is it Like and How Might We Support it? In Proceedings of *the ACM Conference on Human Factors in Computing Systems (CHI 1994)*, pp. 131-137. ACM Press.

Witten, I. H., & Frank, E. (1999). *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*: Morgan Kaufmann.

Zacks, J., Tversky, B., & Iyer, G. (2001). Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology: General, 130*(1), 29-58.

Zhang, D., & Lee, W. S. (2003). Question classification using support vector machines. In Proceedings of *Special Interest Group on Information Retrieval (SIGIR 2003)*, pp. 26-32. ACM Press.

Zijlstra, F. R. H., Roe, R. A., Leonova A.B., & Krediet, I. (1999). Temporal factors in mental work: Effects of interrupted activities. *, Journal of Occupational and Organizational Psychology, 72*(2), 163-185.