

Quality Versus Quantity: E-Mail-Centric Task Management and Its Relation With Overload

**Victoria Bellotti, Nicolas Ducheneaut, Mark
Howard, Ian Smith, and Rebecca E. Grinter**
Palo Alto Research Center

ABSTRACT

It is widely acknowledged that many professionals suffer from “e-mail overload.” This article presents findings from in-depth fieldwork that examined this phenomenon, uncovering six key challenges of task management in e-mail. Analysis of qualitative and quantitative data suggests that it is not simply the quantity but also the col-

Victoria Bellotti is a social scientist with an interest in computer-mediated communication, computer-supported collaborative work, and task management; she is a Principal Scientist and area manager in the Computer Science Lab (CSL) at the Palo Alto Research Center (PARC). **Nicolas Ducheneaut** is a multidisciplinary scientist, with interests in computer-supported cooperative work, computer-mediated communication, and the sociology of online communities; he is a member of the research staff in CSL at PARC. **Mark Howard** was a computer scientist with an interest in the development of experimental systems, artificial intelligence, and extreme programming; he was also a member of the research staff in CSL at PARC. **Ian Smith** is a computer scientist with an interest in the integration of software development tools and practices with ethnographic techniques in user interface development; he is a researcher at Intel Research Lab, Seattle. **Rebecca Grinter** is a social scientist with interests in computer-supported cooperative work and empirical methods to examine human action relating to computer systems; she is an Associate Professor at the College of Computing, Georgia Institute of Technology.

CONTENTS

- 1. INTRODUCTION**
 - 2. TWO STUDIES OF E-MAIL COLLABORATION**
 - 2.1. Study 1: E-Mail Use Variations
 - 2.2. Study 2: Task and Project Management in E-mail
 - 3. WHAT IS TASK MANAGEMENT IN THE CONTEXT OF E-MAIL?**
 - 4. WHAT IS INTERDEPENDENT TASK MANAGEMENT IN THE CONTEXT OF E-MAIL?**
 - 5. HOW DOES TASK AND PROJECT MANAGEMENT PLAY OUT IN TERMS OF THE EXCHANGE OF MESSAGES?**
 - 6. WHAT ARE THE IMPACTS OF E-MAIL-CENTRIC TASK AND PROJECT MANAGEMENT ON E-MAIL USERS?**
 - 6.1. Keeping Track: The Impact of Complex Threads
 - 6.2. Inbox Versus Folders: The Impact of "To-Dos"
 - 7. HOW CAN MAIL TOOLS BE IMPROVED TO SUPPORT E-MAIL-CENTRIC TASK AND PROJECT MANAGEMENT?**
 - 7.1. TaskMaster: Developing a New Model for Handling E-Mail
 - 7.2. The TaskMaster Design Philosophy
 - Threaded Task-Centric Collections: Breaking the Messaging System Metaphor
 - Application Neutrality
 - Task-Centric Meta-Information for All Items
 - Aggregations of Information for an Overview
 - 7.3. Implementation of TaskMaster
 - 8. EVALUATING E-MAIL AS TASK MANAGEMENT**
 - 9. FINDINGS**
 - 9.1. Extent of TaskMaster Use
 - 9.2. Experience of E-Mail as Task Management
 - All or Nothing Use
 - Threaded Task-Centric Collections
 - Equality for All Content
 - Task-Centric Meta-Information for All Items
 - Aggregations of Information for an Overview
 - Summary of Findings
 - 10. ONGOING WORK AND CONCLUSIONS**
-

laborative quality of e-mail task and project management that causes this overload. We describe how e-mail becomes especially overwhelming when people use it for tasks that involve participation of others; tasks cannot be completed until a response is obtained and so they are interleaved. Interleaving means that the e-mail user must somehow simultaneously keep track of multiple incomplete tasks, often with the only reminder for each one being an e-mail message somewhere in the inbox or a folder. This and other insights from our fieldwork led us to a new design philosophy for e-mail in which resources for task and project management are em-

bedded directly within an e-mail client as opposed to being added on as separate components of the application. A client, TaskMaster, embodying these ideas, was developed and tested by users in managing their real e-mail over an extended period. The design of the client and results of its evaluation are also reported.

1. INTRODUCTION

Many knowledge workers spend a considerable amount of their computer time using e-mail, and lately it has become clear that it has taken on the role of a key task-management resource (Bellotti & Smith, 2000; Bellotti et al., 2003; Mackay, 1988; Whittaker & Sidner, 1996). Indeed, the main computer tool used to support project management and informal workflow seems to be e-mail for many knowledge workers (Ducheneaut & Bellotti, 2003; Venolia, Dabbish, Cadiz, & Gupta, 2001).

However, there is very little direct and detailed data that directly examine what is going on in terms of the actual messages sent and received, probably due to the sensitivity of the data of interest (Fisher & Moody, 2001, is a notable exception, although at the time they collected data, they were only able to capture messages that were not deleted in the normal course of e-mail use). Instead, researchers have largely relied on interview data where users report on their practices. We feel interviews are an excellent start for understanding the issues and problems but not enough to understand their importance and extent.

In this article, we report the results of extensive research (conducted over 3 years) in which we sought to understand what it means to manage tasks in e-mail. Namely

- We report briefly on a study that compared the ways in which e-mail was used in three organizations, one being our own.
- We report in more depth on the findings of a second study in which a number of people in our own organization were interviewed about their e-mail practices. We also analyzed longitudinal samples of seven employees' e-mail and observed study participants as they read, organized, and acted on incoming messages. Based on this, we suggest that e-mail overload is more than just a matter of quantity. In fact, complexity of e-mail management can be related to the particular quality of an e-mail user's activities. Overload also seems to be a result of the demands of interdependent tasks; those with delays introduced by waiting for responses from others on whom the task depends.
- We present the design and user-centered evaluation of an e-mail client, which offers solutions for the overload problem, particularly supporting the management of interdependent tasks.

Before we move into describing our studies, we begin with a description of previous work in the area of e-mail-based task management.

Prior work in the area of e-mail-based task management is scarce and the best-known field studies have largely characterized task management in e-mail as a matter of categorization or classification, using a small number of seemingly discreet dimensions to easily distinguish e-mail users' practices. For example, Mackay (1988), in an early study of task management, described people as "prioritizers" or "archivers" in how they handle the incoming message deluge. In a more recent example, Whittaker and Sidner (1996) posited that the working information in the inbox contains "to-dos," "to-reads," and "indeterminate status" items that are hard to "deal with." They classify e-mail users as "no-filers," "spring-cleaners," and "frequent-filers" to characterize different filing strategies that are adopted to try to manage the never-ending influx of messages containing requests for action.

On the more design-centric side of the research, filing and categorizing are again at the forefront of the e-mail research agenda: for instance, various researchers, principally from the Information Retrieval community, (e.g., Bälter & Sidner, 2000; Boone, 1998; Mock, 2001; Segal & Kephart, 1999; Takkinen & Shahmehri, 1998, 1999) analyze "to-do" management in terms of a cognitively difficult classification problem. Their solutions all involve some automatic categorization of incoming messages to diminish the user's cognitive burden.

In short, past research on e-mail task management has portrayed e-mail as a filing cabinet: messages are discrete units put on a more or less regular basis into user- or machine-defined "buckets." Not surprisingly, the biggest challenge identified from this type of research is e-mail volume: as more and more messages pile up in the inbox, the time it takes to file or retrieve messages quickly increases (Bälter, 2000) until the user feels overwhelmed.

However, we believe that volume alone does not explain why e-mail overload occurs. In attempting to deal with often onerous task management, we have found that people tend to adopt configurations of tactics that are as varied as the people themselves and tend to be missed in the effort to classify users into simple types (Mackay, 1988; Whittaker & Sidner, 1996). These tactics highlight additional important information about management challenges, beyond managing volume, and deciding whether to file (or not), associated with the management of overload. We can summarize these challenges as being related to the quality rather than the quantity of e-mail and it is this quality that we examine in detail in this article.

2. TWO STUDIES OF E-MAIL COLLABORATION

We conducted our investigation of task management in two parts. First, in the summer of 2000, we gathered background data about e-mail use from

three organizations with an *in situ* interview-based study, to assess variations in e-mail use across knowledge-intensive organizations (Study 1: this study is only reported briefly later). Second, we conducted an in-depth study, monitoring actual e-mail traffic over 2 or more weeks in our own organization, with video observation of e-mail handling sessions backed up by interviews (Study 2: this study is reported in detail in the following subsection).

2.1. Study 1: E-Mail Use Variations

We have long been interested in the practices of knowledge workers who do not follow rigid procedures or use workflow tools, and indeed, over the years of our own research on task management and e-mail, we have found few examples of knowledge workers doing so. Rather, knowledge workers seem to use e-mail as their main resource to support task management on an ad hoc basis. Supporting this claim, in a recent study (Bellotti et al., 2004), we counted 882 “to-dos” across 16 individuals with diverse job descriptions (3 administrative staff members, 3 managers, 5 researchers from our own organization, and 5 people from other organizations: 2 sales managers, 1 retail manager, 1 charity manager, and 1 university professor). Of these, the largest proportion of “to-dos,” 35% were represented in e-mail, which was as many as in “to-do” lists and the calendar combined (the next most popular resources).

To understand the kinds of task-related processes taking place through e-mail, we conducted 28 interviews in three professional organizations; our own research organization (at the time, 400 employees, and part of an 80,000-employee corporation using a variety of e-mail clients), a rapidly-growing multimedia production start-up (150 employees, using Eudora™), and a small design consulting company (six employees, mainly using Microsoft Outlook Express®).

The interviews took place in the interviewees’ offices, where they could show us the contents of their e-mail. In the first part of the interview, we covered background information reported in Ducheneaut and Bellotti (2001b). The latter half of the interview consisted of interviewees being asked to pick the 10 most recent (nonsensitive) e-mail messages they had received, open them, and explain the collaborative activity they were part of, while drawing that activity on paper, using any representational style they chose (Figure 1 is an example of one of the activities that were drawn by our interviewees).

The (approximately 280) activities we collected were coded by two researchers, working independently (the rate of disagreement in assigned codes was negligible). We found five types of activity with 8% being nonactivity related and two being rather rare (surprisingly, coauthoring and document reviewing was one of these rare types). The frequency of each type is listed in Figure 2.

Figure 1. An example of an e-mail activity sketch.

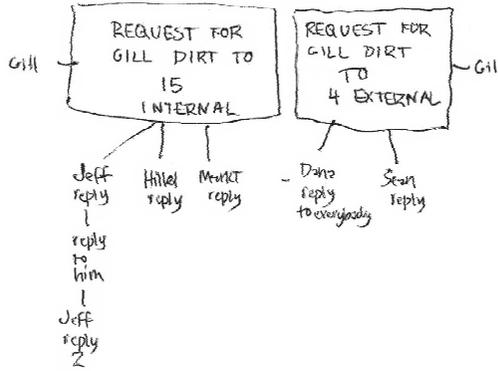


Figure 2. Percentages of Activities Matching Coding Schemes Covering Type of Activity

Activities Coded As	Examples and Notes	Percentage
Announcements	For example, agenda, out-of-office	35
Dialogue, discussion, negotiation	Includes questions	28
Organizing, arranging, coordinating, scheduling	Includes assignment of responsibility	27
Not an activity	For example, spam, out-of-office auto-reply, bounced messages; this low proportion is a result of the use of spam filters	8
Coauthoring, document review	Includes paper-writing and graphics design	2
Formal information gathering	For example, form filling	1

Due to our participants’ inability to enumerate people beyond 20, we were unable to get a precise mean and standard deviation for the number of participants (people who received or sent messages in an activity) but the median number of people involved in these activities was 6, with

- 30% of activities involving only two people
- 10% of activities involving three people.
- 5% of activities involving four people.
- 4% of activities involving five people.
- 45% of activities involving more than 20 people.

We also coded for how many steps and people there were in the whole activity. Including single message activities, we obtained an average length of 3.1 steps (standard deviation 3.6), with a median of 2 steps.

From this analysis, we were able to deduce that our own organization only differs significantly from these other two organizations in how messages are typically addressed.¹ At our organization, fewer people are addressed individually in a multirecipient message to an ephemeral group of people, and more people are addressed as members of a stable group or as a result of belonging to the entire organization (e.g., “**everyone@parc.com**”). This reflects the stable teams arrangement at our (longer-lived than the other two) organization and its belonging to a large company with a greater prevalence of corporate announcements, together with a, unique in this study, readily accessible Web interface for nontechnical users to set up distribution lists. In other aspects, however, the similarity was striking. Consequently, we were able to assume that people in our organization use e-mail in ways that make them fairly representative of knowledge-work professionals.

2.2. Study 2: Task and Project Management in E-mail

During 2001, we conducted extensive design work toward supporting task management in the e-mail channel, leaving the question of the design of clients alone, which is reported elsewhere (Bellotti, Ducheneaut, Howard, & Smith, 2002). However, in 2002, we returned to the question of whether the myriad features in popular e-mail clients do actually provide appropriate resources to manage the kinds of task and project activities we saw in the first study reported earlier. As we shall see, it soon became apparent that there was a particular kind of e-mail-based task (complex tasks involving dependencies on others) that seemed to be causing the most trouble for task management with existing tools.

We conducted an in-depth study in our own organization to examine task and project management in e-mail from a variety of e-mail users' information management perspectives. We were also looking for ways in which current mail tools can be improved on with respect to this phenomenon. There were several phases to this study.

We began with a preliminary survey of mail tool use. The survey indicated that even our technologically sophisticated population only uses a fraction of the features of Microsoft Outlook[®], the dominant e-mail tool in our organization. Further exploratory interviews within and outside our own organization and the revisiting of transcripts and notes from previous interviews from our

1. We used Fischer's exact tests and any significance was obtained at $p \leq 1.5\%$.

studies over the years (totaling over 40 interviews) showed that managers are the most likely to feel overloaded and unable to track the many concurrent processes for which they are responsible. These findings led us to develop a set of six key challenges of task management in e-mail (we do not claim that these are a complete set for all people). We itemize these later and, by way of evidence for each one, we provide quotes of the sort that led to our deriving these challenges (note: some of these are paraphrasings taken from detailed interview notes, rather than transcripts):

Challenge 1: Keeping track of lots of concurrent actions

- a. My own actions.
- b. Outstanding actions (things other people are doing for me).

Project Manager: “Often I don’t send out that weekly status report which could affect customer satisfaction.”

Director: “I have to remember that something is actually needed by someone. Then I have to find it and remember to send it. Then I’ve got to remember that I sent it and I’m waiting for a response”

Principal Scientist: “In project reviews there is a big risk of forgetting to make sure something happens, like the reviewers get a thank you note in e-mail. I have to remember to make sure they send me their slides, and then I have to put them on the web page. There’s lots of details to forget.”

Corporate Strategy Manager: “There’s a long thread of communications back and forth about the action items. As you can see in the example, they go back a few months. One of the problems with these reports is that you have to trust that others implicated in an action are copied by the owner of the action in a timely way, not an hour or a half before the next meeting. So, we need ongoing feedback about whether someone is acting on an item before the next monthly meeting.”

Challenge 2: Marking things as important or outstanding

CEO: “I’ll maybe look back at flags for a month [...] I’m trying to work it down to where my flags really mean something. [...] I’m constantly trying to eliminate them, but I can’t. [...] I did try using the grayed flags feature to indicate actions done, but it wasn’t worth it.”

Director: “I scan who [e-mail is] from and open it if it’s someone or something important. Whether someone else marks the message as important is generally useless to me. I use my own flags to mark things as important enough to be followed up. [...] I try to respond immediately if I can, but at the end of each day, I look at all the flagged messages and

then drag them into my Outlook to-do list and set a deadline for each one. [...] I like to folder the important stuff in contextualizing folders...” [by which he means that he’s got everything relating to one context together].

Challenge 3: Managing activity extending over time or keeping track of threads of activity

CEO: “I was trying to arrange lunch with a friend and then I lost track of the thread, so I forgot to send a reply when we were coordinating. [...] He got rather offended.”

Director: “You are so busy making decisions and paying attention to the here and now; making sure you at least dodge the knives, that you never have time to take a longer view and get a sense of continuity. You just try to sort by date, sender or subject to bunch stuff that’s related together.”

Challenge 4: Managing deadlines and reminders

Public Relations Manager: “I have so many follow-up reminders on due dates for things, and I do this manually for now... This is how I do it now: I have a (paper) folder. This person is going to be here for an interview, and I want [Person1] and [Person 2] to see him but they were not here. So I left a voice mail, then I sent an e-mail, didn’t hear anything, so I left another voice mail because I needed to know right away... But if I get thirty e-mails per day, in 3 days I have another 90 e-mails that I am going through and I have completely forgotten to follow-up on this, unless this (the folder) is visually up.”

CEO: “I have to put to-dos in the calendar, [...] because I have such a tight schedule that I need to block-out time specifically for getting things done otherwise I never will.”

Challenge 5a: Collation of related items including e-mail and documents (e.g., discussions about work content or responses to a survey or set of questions)

Manager: “We have this project called [Project Name], with all the same people. And we switch between one business and the other constantly, during meetings and on-line. I have a [Name1] folder and a [Name2] folder, and I have no idea where to put the messages.”

Principal Engineer: “I do not want to receive all my files through e-mail [...] e-mail should be kept together with your files, or more that

everything should be synchronized with documents stored elsewhere...”

Research Scientist: “here is a set of messages that are related to a topic, but the topic is emergent. The right place to put these messages is not yet clear, I could put them in a folder somewhere and then rename it, but it gets to be a pain if you have to deal with hundreds of messages. So it’s capturing structures as they emerge.”

Graphic Designer and Production Artist: “I try to organize things, and I over organize. So then I try to pull back, I revisit every 6 months and say no, I don’t need that. It’s like clothing you know, if you don’t wear them in a year you throw it away. I want to make things simpler. But you don’t know what to expect when you first start something, so you always over-organize, and then you pull back.”

Challenge 5b: Event-based collation of documents and discussions (i.e., meetings are a salient organizing principle for collections)

Director: “In our [Customer] board meeting recently, there were a lot of presentations to be given and issues to be discussed. The meeting required a lot of organization ahead of time. There were some operational issues, such as who would be there, who needed copies of what things in advance, seating arrangements, dealing with [foreign] cultural expectations [...] all in addition to preparing the content, material for the meeting. The event is the focus of all this activity. [...] It would also be nice to be able to relate it to previous events, for example to get hold of the latest versions of research presentations.”

CEO: “I would really like to be able to attach threads and their related attachments to calendar events.”

Challenge 6: Getting a task oriented overview, at a glance, rather than scrolling around inspecting folders

Corporate Strategy Manager: “It’s hard to recognize projects. There is this extensive repository of project activity. A lot of messaging and inspection of status of things could be improved. Even automating the reminders and the tracking of actions would be helpful.”

Senior VP of Marketing: “The worst I have ever gotten was 400 backed up, which makes me crazy because I know I am forgetting something in there.”

To understand the details of these challenges in terms of what it is that people have to deal with in their e-mail, we recruited 11 participants for a longitu-

Figure 3. Roles and Participation in Study Phases

Role	Study Phase Participation			
	1. Capture Messages	2. Preliminary Interview	3. Video of Triage	4. Follow-Up Interview
Attorney (A)	—	X	X	—
Legal Secretary (LS)	—	X	X	—
Business Development (BD1)	—	X	X	—
Public Relations (PR)	X	X	X	X
Group Manager 1 (M1)—E	X	X	X	X
Group Manager 2 (M2)—E	X	X	X	X
Group Manager 3 (M3)	—	X	X	—
Intellectual Property (IP)	X	X	X	X
Researcher 1 (R1)	X	X	X	X
Researcher 2 (R2)	X	X	X	X
Researcher 3 (R3)	X	X	X	X

Note. Four participants did not permit message capture due to legal or confidentiality concerns. All are Outlook Exchange® Users Apart From Those Marked ‘—E’ Who Use Eudora®.

dinal study of e-mail practices, although not all were able to agree to all the phases (see Figure 3). Indeed, our methods were necessarily onerous for the participants, and many people who complained about e-mail overload felt unable to take on the time commitment and the openness about e-mail content that was required. The four phases of the study were as follows:

Phase 1

For seven participants, we installed e-mail filters for a number of weeks to capture all of their incoming and outgoing e-mail (we did not suffer from the routine e-mail deletion problem experienced by Fisher & Moody, 2001). A special “signature file” template was created and customized by participants to warn correspondents that the participant’s e-mail would most likely be read by researchers. However, participants were instructed to delete confidential or personal content from the folders where the filters stored data (participants reported deleting about 10% of the messages). The filtered data was subsequently transferred to spreadsheets and 2 to 3 weeks worth of data in the middle of the time period monitored was analyzed message-by-message in the following terms:

- *Addressing*—Whether each message was individual, multiple, or list addressed (implying its likely import in terms of thread tracking—e.g.,

Bälter and Sidner, 2000, argued that the fewer recipients there are of a message, the more important it is).

- *Threading patterns*—Series of messages, topically-related replies, and additional related messages (impossible to analyze automatically due to changing subject lines and participants). In this case, threads taking place on distribution lists were only included if the participant in the study actively contributed to them (we assumed that the participant was not engaged with a thread to which they did not contribute).

Phase 2

After filters were installed, we conducted observations of 20 to 60 min of e-mail “triage.” Venolia et al. (2001) defined triage as the practice of handling incoming e-mail and related “to-dos” and organization of messages and attachments. We arranged to be in the participant’s office at a normal time for them to be engaged in this activity (usually first thing in the morning or perhaps after lunch). If there was less than 20 min of triage, we returned for a second session. Participants were asked to talk about what they were doing, but to try to work, as much as possible, as normal.

Phase 3

These observations were followed a few days later by a first round of interviews, covering the participant’s role and use of e-mail. We asked open-ended questions about certain issues (e.g., “what is your biggest organizational headache in e-mail?”) and used 5-point Likert-type scales to get participants to rate how much or often they did certain things or used certain features of their e-mail client.

Phase 4

At the end of our filtered data collection period, we conducted follow-up interviews with the seven participants who permitted this to check our analysis for accuracy of interpretation of the activity and to cover task and document management associated with e-mail threading. The analytic check involved picking three threads “touched” on a day in the middle of the data collection period and stepping through the messages to get the participant’s perspective on the activity (much like the analysis in our earlier study but using real e-mail data, including thread information, rather than recollection as data). We only found 2 cases out of 21 that showed analytic misinterpretation. These were minor; one involved our recognizing threaded messages correctly but combining separate threads about one topic into a single thread,

which the participant saw as separate but related threads; another involved accidentally adding a single message to a group of people on one thread to another thread between the same people at the same time. Otherwise, our analysis was often even more accurate than people's recollections of what took place in the thread (initially someone would occasionally disagree and then, after reexamining the messages, would decide that we were correct). Thus, we have a fairly high level of confidence in our analysis.

In the following sections, we focus on some of our findings about e-mail-centric task and project management from this study. However, we want to avoid putting people into categories as a way of talking about their practices. This has occurred in prior work (e.g., Mackay, 1988; Whittaker & Sidner, 1996) but we feel it is slightly misleading, albeit illustrative. So we use the following terms throughout the rest of this article to characterize phenomena governing practices at different levels of abstraction, rather than to classify people.

Objectives

Knowledge workers must accomplish a number of simultaneous goals with their e-mail such as keeping up to date, responding to requests, tracking requests to others, making deadlines, managing information as it comes in, distributing information, and being prepared for events.

Strategies

To accomplish objectives, we observed a number of strategies to, among other things:

- Keep relevant content at hand.
- Preserve the ongoing work-state of incomplete activities.
- Save content that might be needed again in the future.
- Remind themselves of things to do in the future.
- Find things in the overwhelming and generally growing mass of content.
- Prioritize the “must-dos” against the “would-be-nice-to-dos.”
- Get rid of irrelevant content, and so on.

Tactics

The strategies discussed earlier are accomplished using particular tactics that depend on the person and the affordances of the resources they have. Knowledge workers use information management resources that are easily available within the current active work environment for executing their strat-

egies (Bellotti & Smith, 2000). These resource-dependent tactics seem to vary widely but serve common purposes at the strategic level. In Bellotti et al. (2004), we found that people in our own organization kept two thirds of the “to-dos” they had represented online in e-mail (this is excluding physically represented “to-dos”). In e-mail, task management tactics that we have observed include the following:

- Regularly scanning the inbox; often scrolling up and down.
- Sorting, largely by sender or by flags, to find items more easily than in the default time-and-date-based view.
- Deleting items to clean-out irrelevant, distracting content in the inbox.
- Storing currently relevant items on the Microsoft Windows® task bar or minimized in the Eudora® application window.
- Marking e-mail messages as unread to remind themselves to look at them again.
- Storing items in appropriately labeled e-mail folders and subfolders to be worked on together in the future.
- Archiving messages in e-mail folders for possible future reference.
- Saving attachments in regular system folders for reference.
- Inspecting or searching in folders in e-mail and elsewhere.
- Making a calendar event to remind oneself to do something, and checking and marking the calendar to confirm availability for proposed activities sent in e-mail before responding.

At the tactical level, we see an enormous amount of variation that reflects user exploitation of the complexity and particular characteristics of modern e-mail tools and other computer resources. For example, Eudora users with PCs seem to use the property of being able to collapse message windows in the application window into miniaturized message window proxies to keep track of active “to-dos.”

Outside of e-mail, we also see varied use of paper-based resources such as sticky notes, day-timers, notepads, and printed e-mail messages, in combination with computer-based resources such as the calendar or spreadsheets. Although tactics tend to be what we directly observe, it is the strategies they serve that are analytically more interesting. From a design perspective, and more generally speaking, a user-centered design approach of trying to support every possible tactic may only lead to “feature creep.” It is easy for the analyst to get bogged down studying the infinitely variable (and often fascinating) practices of e-mail users and to lose sight of the larger issues (objectives and strategies) that led to the adoption of these tactics in the first place.

Given that these phenomena suggest how people are using their current tools to manage tasks in e-mail, we now address in more detail some concerns relating to e-mail-centric task management:

- What is task management in the context of e-mail?
- What is interdependent task management in this context?
- How does task and project management play out in terms of the exchange of messages?
- What are its impacts on e-mail users?
- How can mail tools be improved to support e-mail-centric task and project management?

In Sections 3 to 6, we look at three classes of tasks conducted largely in e-mail: rapid-response, extended-response, and interdependent tasks (the first two depending only on the one person and the last depending on the action of more than one person). We argue that, although they are in the minority, it is the interdependent tasks that impose the greatest burdens for task management because they involve managing and tracking the actions of others with little support from current commercial e-mail tools. In Section 7, we turn to our design efforts to combat e-mail overload as we have come to understand it.

3. WHAT IS TASK MANAGEMENT IN THE CONTEXT OF E-MAIL?

According to Merriam-Webster's dictionary, multitasking refers to a *computer* that is capable of performing multiple jobs simultaneously. So it is ironic, if not surprising, that computer use has increased rather than decreased the number of tasks that e-mail users are expected to manage at any one time. In a recent "snapshot study," where we counted all the tasks participants were tracking (Bellotti et al., 2004), we found eight employees in our organization keeping track of 74.4 active "to-dos" (median 65) using, on average, 11.25 different kinds of resources (median 11.5), including calendars, lists in notebooks, and jottings on slips of paper, with e-mail being by far the single most important of these. This "overload" reflects the increasing speed at which tasks can be handled with computing resources, and the increased networked connections between people, increasing the possibility and ease of making requests and demands of others (Sproull & Kiesler, 1991). And, as Mackay (1988) put it, "it should not be surprising that the effect of lowering the cost of delegating tasks by e-mail increases the volume of e-mail" (p. 351).

Of course, busy people, unlike computers, are unable to do many things simultaneously. Instead, they switch among separate activities with different

objectives. This phenomenon has been observed by other researchers to be a common practice, particularly for senior managers. For example, Mintzberg (1973) talked about chief executives' work as being characterized by "brevity, variety and fragmentation" (p. 31), with each deskwork activity lasting, on average, 15 min and each phone call 6 min. Sproull (1984) estimated that managers work on 58 activities per day lasting, on average, 9 min (including, on average, three 1-hr meetings, suggesting many activities take less time than 9 min), and Reder and Schwab (1990) also observed senior managers engaging in 30 distinct tasks per day.

Our video observations of people handling e-mail during triage sessions show that for both managers and nonmanagers, incoming and saved e-mail represents a series of different tasks which are also interleaved. Within a given 10-min interval of videotaped work, many separate tasks would often be handled (we do not have exact times for tasks as it was sometimes difficult to ascertain where one task ended and another began), as our participants switched attention from one incoming message to another. For most of our study participants, some of these tasks required only a simple rapid response, taking as little as a few seconds. For example, a canonical pattern (taken from BD1, a business development specialist, but not intended to characterize all her work) tends to look like this example from our video recordings of e-mail triage:

3.1. An Example of Three Rapid-Response Tasks Within 5 Min

- T1: Open message about a technology development project: follow link in web browser and inspect site about the technology.
 Make to-do note in paper pad.
 Switch back to message.
 Switch back to web browser to inspect site.
 Switch to calendar and check dates.
 Return to e-mail message and open, write and then send reply.
 Close and file message in folder
- T2: Look at next message in inbox.
 Switch to web browser and search for sender in Google to find out who he is: return to inbox.
- T3: Move to next message.
 Use information supplied in message to update distribution list name in online contacts list.

This series of three activities, each prompted by the opening of a new unread message in the inbox and each relating to different subjects and involving switching between applications or different components of Microsoft

Outlook, takes place in under 5 min. These activities reflect some of BD1's key objectives, which are to respond to requests, to keep informed about project information and the groups that do the project work in our organization and follow up on new contacts, business leads, and business development activities. So one of BD1's strategies is to be highly responsive to all incoming announcements and queries in her e-mail. Each of the messages in the example discussed earlier causes some activity to take place outside of the e-mail viewer, requiring a lot of window management and attention shifting. In T1, BD1 learns about a technology project that she needs to work on and confirms that she can make a meeting. In T2, she finds out about a new contact. In T3, she updates a distribution list name in her contacts list. However, none of the activity in this example requires BD1 to communicate with anyone else other than ending the task with a response to a message sender (if need be). The work is nonproblematic from a task management point of view, in the sense that it is easy to execute and then forget about, using the resources at her fingertips.

However, one problem that we do notice in this work is window management. A good deal of opening, closing, and arranging of windows occurs, which, although not extremely time consuming, seems to be inconvenient. In our later discussion of our design efforts, we add this factor to our other task management challenges as something that can be addressed by better task-centered design of e-mail clients. However, Czerwinski et al. (2003) have also demonstrated that having greater screen real estate can also reduce this type of problem.

However, things are not always so simple as is the case with rapid-response tasks. Greater complexity is illustrated in an example of another type of task, obtained during our observation of a public relations manager (PR) handling her e-mail. We saw her constantly returning to a single task, which was to explain how our organization could participate in a TV program about a certain research area. The pattern (the task was started but not completed during our video observation) was as follows:

3.2. An Example of an Extended-Response Task (Incomplete After 1 Hr of Videotaping)

Read message requesting participation in TV program.

Open reply window and begin to list researchers and ongoing relevant projects.

Switch to web-browser and consult web-based phone list for names and spelling.

Consult internal web pages for project descriptions and researchers.

- Consult online web-based document archives and download, open and skim research reports.
- Add names and descriptions based on these searches to reply.
- Collapse reply to task bar without sending.
- Switch to responding to another incoming message in inbox.
- Delete incoming message.
- Reopen reply.
- Consult web pages.
- Add notes to reply text: etc. ...

As in the previous example of rapid-response multitasking, there is a lot of application switching going on here (mainly between Microsoft Internet Explorer®, Adobe Acrobat Reader™, and Microsoft Outlook). But the main activity described here takes so long that it is constantly suspended and the reply is saved on the task bar several times while PR executes related and unrelated tasks in her mail tool and other applications. Because she knows she will probably get interrupted from her current task, she puts in notes to remind herself of things to include in the reply message she is creating. She explains as follows:

Those are kind of the holes that I'm going to need to fill in, in case... it's three o'clock now... in case 4:30 rolls around or some other emergency happens and I don't get to finish and it helps me, so I don't lose my chain.

This more extended category of tasks can demand a strategy for saving a lot of half-completed e-mail responses, each of which includes notes about things to include. PR has two tactics for dealing with this: one is to collapse many open items onto her Microsoft Windows® task bar and another is to save drafts in her drafts folder in Microsoft Outlook. The latter tactic was adopted after a period where her machine was crashing a lot, causing her to repeatedly lose items on the task bar, but now that her machine is more stable she has switched back to the task-bar solution (this tactical switch both suggests that the drafts folder is clearly less convenient than the task bar and highlights how susceptible tactics are to variations in particular circumstances, although the same strategic needs are met).

Thus we have two kinds of tasks prompted by e-mail, with only one requiring significant task management resources.

Rapid-Response Tasks

These are obligations that can be dispatched quickly in a “fire-and-forget” fashion using available resources on one’s machine or on the intranet or Internet.

Extended-Response Tasks

These are obligations that take time to handle, requiring extra task management strategies in terms of preservation of ongoing work status, possibly with a need to make “to-do” notes so that ideas on how to handle the task are not forgotten. Extended responses are often postponed for a while or are interleaved with other simpler activities, sometimes due to the need to be responsive to higher priority rapid tasks, to clear the decks, or perhaps due to the desire to handle something less stressful for a while. As R2 put it

I take lots of breaks when I am doing concentrated or focused tasks, then I'll take a lot of breaks during things like that. [...] If it's a long reply or needs planning then I would probably go ahead and defer it and if it was important and I had to do it soon then I would probably write a card for it.

Of course, there are no clear-cut distinctions between rapid- and extended-response activities (for example, we saw BD1 make a “to-do” item based on one of her tasks). Indeed, sometimes the rapid responses are part of a larger pattern of extended collaborative work. The distinction we are making here is more properly a matter of the degree to which something can be handled quickly and then forgotten about (perhaps until, in some cases, another message regarding the same activity shows up). Many activities sit in the middle of this time investment and management dimension.

However, neither rapid-response nor extended-response tasks are, we believe, the source of the biggest headache in managing “e-mail overload” (as defined in Whittaker and Sidner, 1996). We believe that a significant source of overloading is an overlooked factor that we call *interdependent task management*.

4. WHAT IS INTERDEPENDENT TASK MANAGEMENT IN THE CONTEXT OF E-MAIL?

Much of the activity of e-mail triage is nonproblematic (albeit time consuming). Putting it simply: messages are received and the recipient composes a response or a person simply sends a new message. Composition time varies widely, using few or many resources at hand. The task is pretty much under that recipient's control. Should a person become overloaded by this kind of task management, it would simply be the quantity of e-mail that would be the major overloading factor.

Other tasks for which e-mail users are accountable, however, are ones that not only imply a need for rapid and extended responses but also depend on someone else to complete an action that a response demands. Of course, that

“someone else” may not be accountable for the task’s completion in the same way that the task owner is. For example, that “someone else” may simply be very busy and distracted with other tasks, he or she may have other priorities in general, or he or she may not be answerable to the task owner in any formal sense. Even so, it is up to the task owner to “manage” the activity of the other person (or people) to complete the task.

All knowledge workers experience this kind of dependency on others across, up, and down the organization to get their work done and e-mail has done nothing if not increase the capacity for these dependencies by increasing the connections between its users within and between particular organizations (Sproull & Keisler, 1991).

An example of this kind of interdependent-response e-mail can be drawn from group manager 1 (M1), who worked on the task example, following, in between other tasks while processing her e-mail:

4.1. An Example of an Interdependent Task

In this example, we include think-aloud comments of M1, which explain the rationale for her actions² (quotes are in plain text and our explanatory notes are in italics):

“I see I’ve got a bunch of stuff on ...TechCo [*a technology corporation*]... I’m going to collect that stuff together, Rob [*M1’s lab manager*] wants me to do something with it.”

One of the messages mentioned as “that stuff” is from Ralph [senior internal attorney]. It’s already been read. M1 looks at her little notepad containing notes and to-do items.

“They’re trying to work on an intellectual property agreement *between [our organization]* and TechCo... and I got a phone message from Marilyn [*internal attorney*] about this, this morning.” She notes the message in her notebook, “I’m going to forward this message to Ralph to Rob.”

M1 creates a reply to Ralph’s message in which he said that Marilyn is the best person to handle the agreement. In her reply M1 tells Ralph that she got a phone-call from Marilyn with a question about existing agreements. M1 copies Rob on her reply presumably to apprise him of the fact that the work is being executed.

“Basically I’m asking Ralph to help me on this because I don’t know what’s already been done on these agreements. Marilyn was asking me

2. We have invented names to maintain confidentiality of our participants.

what's been done and I think Ralph is the only one that knows that's been done..." *M1 sends the reply and then goes back to Ralph's original message in her inbox.*

"Check that off," *she marks in her paper notepad,* "and file this somewhere where I'll be able to find it like under [*M1's lab's name*]. I don't have a file yet for TechCo so I'll create one..." *She goes back to her inbox, sorts on subject line, and searches it and files a couple of messages on this topic.* "Thing is I knew there was more messages on this topic, but they must have changed the subject line, so..."

This is a good example of the kind of interdependency that can occur in a thread of activity occurring predominantly in e-mail (but also via other media like the telephone). M1 has been asked to handle something where she depends on others for information and action and they also depend on her in the same way. The activity is not fully under her control and she must manage the activities of others to accomplish her goal. Interdependency of this type occurs regardless of routineness and formality of an activity, for example, it was very apparent in M1's routine performance-appraisal exercise, involving collating lab members' comments about their peer's activities; in the formal process managed by intellectual property (IP), involving numbering and processing invention proposals, and transfer of formal documents; in M2's discussion and exchange of drafts of slides with a team of colleagues to drive several presentations by the team members to our organization about its own culture; and in researcher 1's (R1) matching of the internship program applicants with possible projects and willing supervisors.

Thus, in our analysis of overload, we must account for the following.

Interdependent Tasks

These are obligations that depend on the action of others to be fulfilled. We predict that if the quantity of e-mail were held constant, those with more interdependent tasks would feel more overloaded because, in addition to needing to manage rapid and extended responses, they have to remember to come back to tasks after being able to forget about them when a request for action is issued to someone else. Further, they have to manage other people's activity to complete such tasks. This is a quality not a quantity issue and is examined in more detail later in the analysis of e-mail data.

As a final point in this section, we want to emphasize that e-mail users in our study invested a great deal of time in organizing and searching through their e-mail content to keep track of their tasks. From our video analysis, we see the following breakdown in the amount of time our participants spent on different kinds of activity:

- 23.1% reading e-mail.
- 6.2% scanning inbox.
- 2.4% deleting messages.
- 2% looking for messages.
- 1.1% spent adding attachments.
- 9.5% filing messages.
- 0.8% opening attachments.
- Most of the rest spent writing e-mail and editing documents.

This adds up to about 20% of time spent just organizing or searching through content. From a design perspective, we see this as a challenge that clearly reflects e-mail users' task-centric needs that are implicit in the six challenges of task management in e-mail mentioned in Section 2.2.

5. HOW DOES TASK AND PROJECT MANAGEMENT PLAY OUT IN TERMS OF THE EXCHANGE OF MESSAGES?

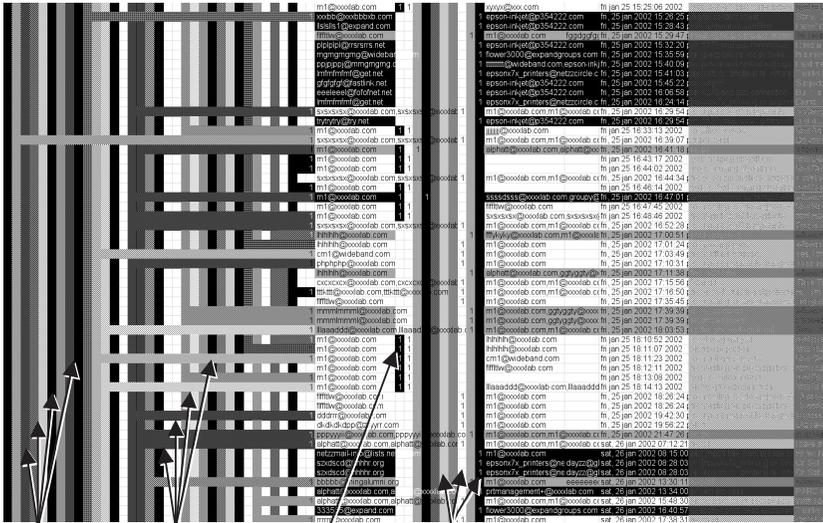
Rapid-response, extended-response, and interdependent tasks are what e-mail task management consists of. But we also want to examine how extensively people conduct task management in e-mail, and understand, to some extent, how much of this task management it takes for people to begin to feel overloaded (as in, e.g., Denning, 1982, and Whittaker & Sidner, 1996). To address this question, we analyzed 2 to 3 weeks of incoming and outgoing e-mail for seven of our participants (those that took part in Phases 2 and 3 of our study):

Three researchers: R1, in addition to her research activities at the time of our study, was an at-large member of our lab's management committee and was managing a student internship program and appeared to have by far the most management responsibility of the three researchers; R2 was coleading a five-person project; R3 was taking over the management of a weekly lab meeting (inviting, negotiating with, and scheduling presenters).

Two process managers: the Public Relations manager (PR) fields outside enquiries about the organization and manages press and publicity. The intellectual property manager (IP) controls the formal process of selection and development of original inventions into patent submissions.

Two group managers: M1 manages two groups totaling 10 researchers and manages a lab-wide budget for over 35 people; M2 manages a single group of 6 researchers.

Figure 4. A sample of the analysis of threads of e-mail (vertical bars on the left) taken from our most overloaded participant, M1. Each row represents data from a single message from a Friday at 3:25 p.m. until the end of the following Sunday. The image has been deliberately degraded to protect the confidentiality of the senders and recipients of messages.



Vertical bars of a given shade indicate threads. Horizontal bars of that line is a member of same-shaded thread. Black cells in this column indicate outgoing. Various other codes in these columns. Each row indicates a message. White rows are those addressed to this person, grey rows are addressed to more than one recipient and black rows are addressed to distribution lists. The shaded areas

Figure 4 shows a sample of the analysis conducted on the e-mail of our volunteers. The filtered and saved e-mail messages (representing all the nonsensitive and confidential e-mail of each participant; roughly 90% of all messages incoming and outgoing) were exported from participants' e-mail filters into a spreadsheet where each message was represented by a single row of cells. In the spreadsheet, all the messages incoming and outgoing were ordered by date to allow us to then read each message and then code it for the following characteristics (among other others which we do not discuss here):

1. Incoming or outgoing.
2. Sent to a single recipient; sent to multiple, named recipients; or sent to a distribution list.
3. A stand-alone item with no follow-up, or part of a thread.

In the figure, the sender address appears as the first cell containing text in each row. This is followed to the right by coding cells in color-coded columns, then the recipient "To:" field, and then the "Cc:" fields appear in the next two cells. These are followed by the date, subject header, and the body, each in their own cell (the cells are not wide enough to see all of the text in many cases and, where nothing is entered in the next cells to the right, the text overflows over those cells). All of the messages that we collected, that were sent and received by M1 from a Friday at 3:25 p.m. until the end of the following Sunday, are shown in the figure. Each row was color-coded to highlight whether it had a single recipient (white); multiple, named recipients (grey); or was sent to a distribution list (black).

Patterns quickly appeared in the data once it was prepared in this way. In Figure 4, it is clear that incoming e-mail was significantly greater (an average ratio of 3:1) than outgoing. Outgoing is indicated in the figure by clusters of messages highlighted with color-coded black cells in a column that is otherwise white, left-of-center. Much more incoming than outgoing mail was addressed to distribution lists (as one would expect), mainly appearing in clusters and clumps. Outgoing messages, mainly one-to-one and also appearing in clusters, tend to be followed by incoming one-to-one and one-to-multiple-recipients messages.

On the left side of Figure 4, vertical bars with horizontal lines branching off represent color-coded threads (shown here in grey-scale). The horizontal bars correspond to rows containing messages that belong to a thread. The analyst started from the cells to the left of the cells containing the sender address, and, by examining the senders, recipients, subjects, and bodies of messages, filled in the columns with threads, and worked outward, creating new vertical bars as needed. All threads noted during the sampled period were traced backward and forward as far as possible beyond that period (sometimes a couple of months worth of filtered e-mail was available). This representation allowed threads to be clearly visualized and counted (e.g., the number of active threads on any given day was clearly visible). The representation also shows very clearly how complicated the interleaving of threads can get with participants often actively engaged in many simultaneous lengthy threads, or engaged in short intense threads.

In Figure 4, M1 engages in 30 simultaneous mainly long threads, only two of which are started and two of which are terminated during the 3 days shown here. This is an exceptional number and reflects why M1 reported feeling it was difficult to keep track. However, the average number of threads participants were engaged with on any given weekday was still quite high (14.8; see Figure 5).

Different participants had different patterns of threading. Some tended to have longer threads, such as R3 (average 5 days) and M1 (average 8.8 days).

Figure 5. Summary of Study Participants' E-mail Data

Role	R1	R2	R3	PR	IP	M1	M2	Avg
Messages received per week day	33.4	25.6	14.3	34.1	39.2	83.2	38.1	38.3
Messages sent per week day	12	5.5	4.2	23	13.8	15.6	10.6	12.1
Uniquely addressed messages received per day (1:1)	8.9	2.6	2.0	11.9	5.4	11.4	11	7.6
Uniquely addressed messages sent per day (1:1)	6.6	2.3	2.0	12.5	7.7	8.2	8.7	6.9
Multiply addressed messages received per day (1:N)	3.3	2	2.4	5.5	5.3	6.2	9.6	4.9
Multiply addressed messages sent per day (1:N)	1	0.9	0.6	3.5	1.2	0.8	1.5	1.4
DL addressed messages received per day (1:DL)	13.6	9.1	5.8	11.3	14.1	45.3	15.6	16.4
DL addressed messages sent per day (1:DL)	0.3	0.8	0.3	2.7	0.1	1.5	0.1	0.8
Threads per week day (TD)	13.3	5.7	6.5	18.7	14.6	25.7	19.4	14.8
Steps per thread average (S)	4.0	3.8	4.8	4.3	3.9	7	4.8	4.7
Days per thread average (D)	2.9	2.9	5	2.4	3.1	8.8	3.9	4.1
Metric of tracking difficulty (D/S) × TD	6.9	3.8	6.0	8.9	10.7	30.2	14.3	11.1

Note. AVG = Average scores; DL = Distribution list. The three highest scores for each category of measurement are in bold to show who tended to have more messages in that category. Note that about 10% of the actual messages sent and received were deleted by participants, due to sensitive content.

R3 reported being involved in long-term e-mail discussions with remote colleagues. M1 was busy coordinating some lengthy, complex processes such as managing a large budgeting exercise involving a lot of other people. Others, such as R1, R2, and PR, engaged in shorter threads (averages 2.9, 2.9, and 2.4, respectively). PR was notable for engaging in many very short bursty threads lasting only an hour or so, suggestive that the role of Public Relations, which we observed, requires fast communications and negotiations with staff at the Palo Alto Research Center to provide information for the press in a timely fashion.

Figure 5 summarizes our analysis of the e-mail sent and received by the seven study participants who permitted this data collection. The picture that emerges from e-mail logging and thread-tracking is, although not statistically significant, certainly suggestive.

First, a very small proportion of e-mail is sent to distribution lists (DLs), suggesting that most work conducted by the participants in e-mail takes place in messages addressed to individuals or specific people in groups that are not well enough defined to become lists. Recall, furthermore, that our organization used DLs significantly more than two other organizations we examined (lists are much easier to create here). This suggests a predominance of dynamically created groups that collaborate via e-mail in professional organizations. Hollan and Stornetta (1992) suggested that ephemeral DLs would be a potentially interesting feature to add to e-mail clients. Their intuition is certainly corroborated by our study.

Second, those with more management responsibilities receive more “important” e-mail (in the terms of Bälter & Sidner, 2000); that is, messages addressed to them individually (e.g., R1 received a lot of individually-addressed e-mail from applicants to the internship program she was managing) or as specifically-addressed individuals (e.g., IP and M1 and M2 were addressed specifically on a lot of lab management-committee messages). Important e-mail suggests greater likelihood of needing to be responsive to a thread and those with management responsibilities do indeed send more e-mail. In addition, managers are involved in more threads on a daily basis, although their threads do not apparently involve more steps than those of others. Our data is also roughly consistent with that of Venolia et al. (2001) in that over half of the e-mail we logged seems to be involved in threads of two or more messages. However, Fisher and Moody (2001) found that only a third of the messages they examined were threaded.

Third, we computed a metric of e-mail-based task-tracking difficulty based on our statistical data. Although this is purely a theoretical extrapolation, it seems to correspond to verbal reports of overload discussed in the next section. The argument for this metric is as follows:

- We can assume that threads, and the tasks they relate to, become more difficult to track, as the intervals between messages grow longer. Delays are a key challenge for the interdependent tasks we discussed earlier; the task must be remembered and tracked as long as one is waiting for a response from a collaborator. The difficulty arises because older messages in the thread drift out of the inbox or get filed and thus disappear from view and cease to act as reminders about the task in the inbox. The e-mail user must now rely on memory (and usually, additional action) to retrieve the task. Verbal reports and video data from our participants suggest that people are forgetting or doing extra work to avoid forgetting threads with these large intervals:

M1 “I scroll up and down my inbox several times a day looking for things I might have forgotten.”

M2 “Inbox [to-dos], it tends to be more of a memory exercise [...]. When they get above a certain threshold in number then I am afraid I will forget about them being in my inbox to deal with.” Note that having more items in the inbox pushes some items out of view.

PR “I find something way down on my inbox, I may re-send it to pop back up to top.”

- To compute the average intervals or delays between messages for each person we divide the average number of days (D) per thread by the average number of steps in threads (S) for each participant, giving D/S . This measure allows us to compare people in terms of the average tracking difficulty of their threads.
- If a person’s average thread has a certain level of tracking difficulty due to delays, multiplying that value by the number of threads a person is tracking gives us an overall relative indication of difficulty. Thus, we multiply D/S by the average number of threads per day (TD) for each person to give $D/S \times TD$. This tracking difficulty score is shown on the bottom row of Figure 5.

Note that delays are not a problem, as long as they are not interfering with the completion of a task; one can simply forget about the thread until a response arrives. However, tasks can have deadlines or at least an issue of time-liness (in Bellotti et al., 2004, we found that after 2 weeks, incomplete “to-dos” are unlikely to ever get done, with 68% being done within a week). Thus, it is necessary to keep in mind ongoing interdependent tasks to make sure that they do not extend beyond a reasonable completion time.

This tracking difficulty metric is highest for the managers in our study. Our public relations officer has the middle score and our researchers (who are individual contributors) have the lowest scores. This metric corresponds to ver-

bal reports of overloading (discussed later) and may provide an explanation for why it is that managers tend to complain the most about keeping track of e-mail activity: Their threads are simply more numerous and have a longer average interval between steps, which makes them harder to remember and to keep track of.

6. WHAT ARE THE IMPACTS OF E-MAIL-CENTRIC TASK AND PROJECT MANAGEMENT ON E-MAIL USERS?

6.1 Keeping Track: The Impact of Complex Threads

In this section, we take a more detailed look at some of the qualitative data that corroborate our statistical data discussed earlier and tend to provide support for our metric of e-mail-based task tracking difficulty.

M1, who has the highest tracking difficulty according to the metric introduced earlier, was the person interviewed in Study 1 who complained the most about being overwhelmed by tracking so many threads of activity in e-mail. These threads generally represented activities in which she was waiting for others to do something for her (she had, on average, 25 ongoing threads per day). Qualitatively, her work was certainly the most extreme in terms of such tasks. During the study, she executed many interdependent budget-management tasks and was using e-mail to manage a highly interdependent annual performance appraisal process. In these two processes, she sent a lot of prompting and reminding messages (often resends and frequently with a subject line starting with the word *Urgent*). As mentioned previously, we conducted five additional interviews with high-level managers and entrepreneurs who were engaged in similar processes, but who stated that they were too overwhelmed to take part in our study.

IP and PR (14 and 18 threads per day, respectively) also complained about having difficulty “keeping up” with their e-mail. PR’s biggest “to-do” headache is “relating processes and retrieving content.” However, she feels she keeps on top of it and indeed her tracking difficulty score is lower than IP’s. IP, with fewer messages but a higher tracking difficulty score, complains that he cannot track other people’s actions in the complex IP management processes he manages. When asked how he keeps track of other people’s actions he said, “Some of it would show up in e-mail, it certainly wouldn’t be anywhere else.” In other words, he depends solely on his e-mail for support with this.

Although M1, PR, and IP feel overloaded or overwhelmed, M2 (with 19 threads per day and the second-highest tracking difficulty score) feels heavily loaded but “in control.” We believe this is because she alone uses an unusual strategy that helps keep track of the actions of others by insisting that her staff

use special key words in their subject lines and automating some of the organization of her e-mail by using filters for those key words:

People send me e-mail that automatically gets filtered to [a particular folder] if it has the right thing in the header... so if they send me reports [...] and that works pretty well... actually I'd say about 80–85% of the time [they] send the things they're supposed to... you know with the right subject line. And maybe like the remaining 15% of the time they don't send it, but the hardest thing is that it turns out I picked sort of a wrong... a bad keyword.

M2's, albeit imperfect, solution creates automatic task-centered collections using filtering and illustrates the structuration described in DeSanctis and Poole (1994). By this we mean that properties of the Eudora client prompt changes in the socially-defined processes that leverage them. Eudora opens a new window for each folder that has new e-mail filtered into it. Thus, M2 does not have to worry about where to sort such messages to and, critically from our perspective, is alerted to them by the opening of the windows.

Together, the qualitative and quantitative evidence we have gathered tends to point to a strong effect of interdependency and difficult-to-track-threading on feelings of overload. In other words, beyond quantity, the quality of e-mail seems to be a substantial factor in explaining this phenomenon.

6.2 Inbox Versus Folders: The Impact of “To-Dos”

In the M1 example in Section 4, activity on her highly-interdependent task is followed by filing away the messages relating to it (M1 has a huge number of folders). Why does she do this, when the activity and its “to-dos” are by no means complete?

We find one of the primary activities in e-mail triage to be information management, for example, scanning, organizing, and retrieving behavior, which takes up about 20% of the time (according to our video data) and is constantly interleaved with actual sending and responding to messages. Most of our participants scan new e-mail and get rid of obvious spam or uninteresting items first. Then they may look at the most interesting messages, and gradually move toward deciding which, if any, to attend to. Deletion and filing occur as messages are dismissed, skimmed, read in detail, answered, or batched for more focused work or archived in case of any future need for their content.

Whittaker and Sidner (1996) characterized e-mail users as frequent filers, spring cleaners, and no-filers. Likewise, our participants range from people who file their incoming messages and may have 30 or fewer in their inboxes, to those who do not file and have over 2,000 messages in their inbox. Our participants also vary in how many folders they have, ranging from none to

hundreds, organized in complex and deep hierarchies. Three of our participants (Legal Secretary [LS], BD1, and IP) manage to keep their inbox down to fewer than 50 messages by filing messages and have adopted the tactic of using their inbox as a “to-do” list (as noted in Whittaker & Sidner, 1996).

However, although Whittaker and Sidner (1996) reported that “to-do” folders tend to fail as a task management resource, our study suggests that folders do in fact work as a tactic for “to-do” management for several participants (particularly M1, IP, and R2). Also, R3 uses a Microsoft Outlook feature that lets him view all his messages by sender; another tactic that seems to support the same strategy. These tactics manifest the strategy of wanting to keep related messages together for processing together in task-focused work sessions.

IP keeps his inbox down to fewer than 50 messages, functioning as a “to-do” list, and yet he also has a “to-do” folder that he created to use for intellectual property management actions and now uses daily for this and some other unrelated actions.

However, we also saw at least two frequent filers (M1 and M2) who do not try to empty their inbox (of roughly 2,100 and 925 messages, respectively) but use some folders as “to-do” lists where they file important messages that constitute specialized activities, while leaving other hard-to-classify “to-do” items in the inbox. Although their inbox still functions as a “to-do” list, it is also full of irrelevant messages. This forces them to use additional tactics like routinely scrolling up and down their inbox, flagging messages, or sorting by sender or subject line. Based on such examples, we feel that classification of people into specific categories (e.g., Mackay, 1988; Whittaker & Sidner, 1996) can be a misleading oversimplification of reality.

7. HOW CAN MAIL TOOLS BE IMPROVED TO SUPPORT E-MAIL-CENTRIC TASK AND PROJECT MANAGEMENT?

Our findings have led us to an alternate conceptualization of e-mail overload. Rather than considering e-mail as file cabinet, with quantity management driving design thinking, we also consider the quality of e-mail-centric task and project management, and particularly interdependent tasks, in a new philosophy for e-mail: that of a task- or project-management tool.

The primary problems, particularly for interdependent tasks, that have emerged from our data as mail client requirements, are as follows:

- Managing and switching between multiple applications or components of applications—Each opening, closing, and repositioning of an application window for Web browsing, document viewing, calendar and con-

tacts viewing, and so forth, creates wasted time managing the computer desktop as was illustrated in our example from BD1, discussed earlier.

- Keeping messages within threads of activity together and in view—Currently there are too many concurrent interdependent activities for more overloaded e-mail users such as M1 (see Figure 4) to keep track of them all in the inbox and at the same time make sense of them. However, foldering is time consuming (Bälter, 2000) and putting “to-do” items in folders makes it harder to remember many concurrent threads and their related actions (following the out-of-sight, out-of-mind principle). Notwithstanding M2’s ingenious filtering solution, which seems to work due to the particular properties of Eudora on her PC (but meanwhile contributing to the window management problem), automatic filing via filtering, or any form of automatic categorization for that matter, can also lead to poor solutions: Studies have shown users do not trust the classifiers (Reder & Schwab, 1990) and like to see the messages before moving them anywhere. Such categorization can also defeat the use of the inbox as a general “to-do” list (but see Segal & Kephart, 1999, for an interesting alternative).
- Making resources appropriate to any given thread easier to access from that thread of activity—Current mail tools compartmentalize resources along application lines rather than activity lines with attachment folders, contacts, and calendar features as separate components (Bellotti & Smith, 2000). This problem also extends to an out-of-date messaging-system metaphor for clients, which places drafts of incomplete messages involved in extended-response multitasking, out-of-sight in a separate folder. Users may, for convenience, resort to the risky (in case of crashes) but more accessible task-bar tactic. In either case, the incomplete message is stored out of the context to which it relates.
- Aggregating information about threads of activity—Current mail tools do nothing to support overviews of tasks. Once an e-mail item is foldered away, it is out of sight along with any information that can serve as a reminder of needed action. Further, if messages marked for action (by flagging, marking as unread, or in Eudora only, by color-coding) are left in the inbox, they can add up to a very large number so that these cues become lost among so many others. For example, LS’s inbox of about 40 messages was dominated by 33 flagged messages with no visible connections between those that were related. Generally, our study participants reported keeping track of anywhere from a dozen to about 100 “to-dos” in e-mail that they intend to act on, and this is consistent with our more recent studies of “to-do” management mentioned at the beginning of Section 3.

7.1. **TaskMaster: Developing a New Model for Handling E-Mail**

In the following sections, we show how we arrived at a design for an e-mail client that tackles head on the issues we raised earlier. Combining our six challenges of task management in e-mail with the matter of managing different applications and windows, we now have seven specific strategic problems that the participants in many of our studies of e-mail practices experience with task management in e-mail:

1. Keeping track of lots of concurrent actions: (a) one's own "to-dos" and (b) "to-dos" one expects from others.
2. Marking things as important or outstanding among the less important items.
3. Managing activity extending over time or keeping track of threads of activity and discussions.
4. Managing deadlines and reminders, which may be associated with particular messages or other content.
5. Collating (a) related items (e.g., an extended thread or responses to a survey) and associated files and links and (b) event-based collation of documents and discussions (e.g., meetings are a salient organizing principle).
6. Most important, getting a task oriented overview, at a glance, rather than scrolling around inspecting folders.
7. Application switching and window management.

Most of the solutions discussed in Section 1.2 focus on 5; collation of related incoming items, by topic or by thread. But collation alone does not address the problems of interdependent task management that we have illustrated in detail here. TaskMaster goes much further than previous efforts to address the seven problems discussed earlier by repositioning e-mail as task management, providing resources to reduce the time-consuming work of overloaded multitaskers. Gartner (2001) estimated that business users spend about 49 min each day processing their e-mail. We felt that a tool that ameliorated problems 5, 6, and 7 could save users some of this time at the very least and, by improving task tracking, would also address problems 1 through 4 and ameliorate indirect (and difficult to assess) costs stemming from failing to satisfy one's commitments.

7.2. The TaskMaster Design Philosophy

Above and beyond all else, TaskMaster is intended to break away from the messaging-system metaphor (featuring inbox, outbox, and drafts folders) and to present a strongly-focused task-centric view of e-mail. A number of principles combine to distinguish TaskMaster from an ordinary mail tool. These are described later.

Threaded Task-Centric Collections: Breaking the Messaging-System Metaphor

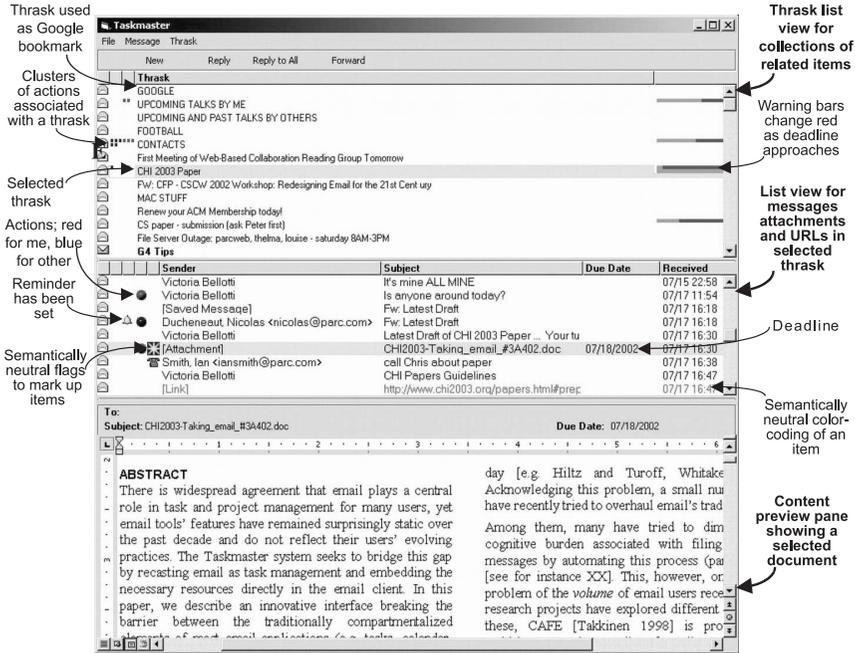
The main element of interest in TaskMaster is the task, not the message. We found that individual messages can represent tasks but interdependent tasks (described earlier) comprise threads of messages files, links, and drafts. So a key principle in TaskMaster is to maintain semiautomatic task-centric collections of mail and related content. We call these collections *thrasks*: part thread- and part task-centric.

In the thrask model, any related incoming messages (replies in a thread, with any attendant files or links) are automatically grouped together by TaskMaster, based on an analysis of message data. Specifically, the application looks for the Exchange server's "reply to" or the Internet standard "in reference to" property of a message, or if neither of these are available, it uses the subject line text. Using this data, it can append a message to the appropriate, matching thrask. This surprisingly simple function³ effectively mimics the filtering tactic of M2 and saves the effort we observed in problem 5 shown earlier, in much the same way as collation systems do.

To avoid problems associated with hidden folders containing new mail and "to-dos," TaskMaster maintains the "to-do" function of thrasks by keeping them in its main list view (the top pane in Figure 6) together with incoming new (non-thrask-related) messages, which appear as bolded until read, single-item thrasks at the bottom of the list, rather like new items in an e-mail tool's inbox. For example, in Figure 6, "G4 Tips" at the bottom of the top pane is a new message that automatically appears as a single-item thrask. This

3. There are clearly problems with this simple-minded approach. Topics in e-mail conversations drift while keeping the same subject line. Further, people will find an old message from a person and reply to it (in some cases with the same subject line!) rather than finding or typing the recipient's e-mail address. So this would seemingly be a recipe for disaster, but in practice, it worked surprisingly well and no users complained or mentioned this "dumb" approach as a weakness. This is an example of where keeping things simple is sometimes the best policy, at least for an experimental research prototype like ours.

Figure 6. The TaskMaster interface—The top pane is the thrask list viewer, the middle pane is the message and other thrask member items list viewer, and the bottom pane is the content preview.



visualization of thrasks and new messages takes seriously the notion of the inbox functioning as a “to-do” list, compressing different related items together under thrasks, which each act as a separate “to-do” item.

The middle pane displays thrask content (messages, attachments, and links). This content can be thought of as a task-centric collection of useful material collected from e-mail or moved or copied from the user’s hard disk. The bottom pane displays a preview of the content of individual selected items (in Figure 6, a thrask entitled “CHI 2003 Paper” is selected and a Microsoft Word® document, a draft paper within that thrask, is being viewed). So thrasks, unlike the folders in most collation systems, remain visible in the main top pane and are optimized as reminders and application-independent repositories for content relating to ongoing tasks, in response to our observing that people often use e-mail folders as secondary, activity-centric “to-do” list collections.

TaskMaster reflects our finding (see problems 1 and 3, shown earlier) that, when managing a task, one’s own drafts and messages are often as important

to keep track of as incoming messages (often representing incomplete tasks as in PR's extended-reponse example shown earlier, or representing "to-dos" for others). So, in a break from the standard e-mail-as-messaging-system model (with inboxes and outboxes and a drafts folder), incoming and outgoing messages and drafts are viewed together. TaskMaster has no outbox, sent, or drafts folders. Outgoing and incomplete messages appear as new thrasks in its top pane only if they are unrelated to existing content. Otherwise messages are added to existing thrasks (as if the sender were copied on every message)—with the user selecting the thrasks in the case of a completely new message or automatically if the message is a reply to existing content in TaskMaster.

Because thrasks correspond to threads of activity not threads of messages, we allow users to rename them to something meaningful as we saw people do with folders (the default is the first message's subject line but this may not be particularly meaningful). We also allow users the flexibility of editing incoming and sent messages as a means of recording notes about activity.

Our analysis (Ducheneaut & Bellotti, 2001b) showed that threads of activity in e-mail do not always correspond to straightforward message threads, so we let users fine-tune the contents of a thrask by adding single items or entire thrasks to other thrasks or by moving items or subthrasks out. In this way, topic-drift in a thread is accommodated (a thrask can be split into multiple thrasks) and technically unrelated threads can be combined. So TaskMaster differs from collation systems that track threads (Cadiz, Dabbish, Gupta, & Venolia, 2001) because thrasks go beyond system-defined threads to encompass user-defined task-centric collections.

Altogether, thrasks principally address the following two problems that we identified:

3. Managing activity extending over time or keeping track of threads of activity and discussions.
5. (a only) Collating related items (e.g., an extended thread or responses to a survey) and associated files and links.

Application Neutrality

Our second principle is *application neutrality*; we do not regard messages as always taking precedence over attachments and links as do classic mail tools. Our fieldwork shows that users frequently cross application boundaries as they work. Messages are often deleted although an attachment must be filed or a link must be bookmarked. Thus, further departing from the messaging-system model, thrasks contain not only messages but also attachments and links as first-class citizens. Attachments and links are broken out of messages as individual items in their own right (see Figure 6) and appear as

other items within the same thrask as the original message that contained them (ideally they should have been visually grouped, possibly by indentation or shading, but this was a refinement we did not have time to make and, surprisingly, no one complained about it). Users can also include items from their desktop or useful links that have never been sent in e-mail.

Just as e-mail message content can quickly be previewed, so can the contents of the other types of items such as Web pages, spreadsheets, presentations, and documents (see Figure 6 where an early draft of this article is being viewed in TaskMaster). All the user has to do is click directly on a document or URL item in the list view to see its contents, reducing tiresome application start-up time and window management (problem 6, discussed earlier). Documents such as Microsoft Word, PowerPoint®, and Microsoft Excel® can also be edited in this pane and users can interact with Web pages by typing in fields and following links. This feature opens up the intriguing possibility of being able to use TaskMaster as a bookmarking tool for favorite URLs, such as our organization's phone list, or Google™ (as in the top thrask in Figure 6).

Application neutrality makes TaskMaster feel less like a classic application and more like a general task management environment, handling a variety of types of media. It mainly addresses the following two problems:

5. (a only) Collating related items (e.g., an extended thread or responses to a survey) and associated files and links.
6. Application switching and window management.

Task-Centric Meta-Information for All Items

Our third principle is that any item in TaskMaster can have task-centric meta-information, such as a deadline, reminder, action, an icon (from a pre-defined selectable set), or a color code (see Figure 6). Even if an item is a document or a link it, might still represent a “to-do,” just as a mail message might.

TaskMaster users can assign meta-information directly to items in two ways. They can click in the areas of the list view where the corresponding meta-information is normally displayed. Or they can click on buttons in an open message window. Users no longer have to copy information into a separate task resource or into their calendar as we have often observed people doing. Deadlines and reminders cause notifications to appear when they are due. Actions are represented as red or blue balls (to represent actions for oneself or another). Color-coding or iconic flags (e.g., the star and the telephone in Figure 6) can be applied to items to distinguish them. However, these resources are semantically neutral to the application itself; they merely make messages more distinctive to users.

The principle of metadata for all items addresses the following problems:

1. Keeping track of lots of concurrent actions—one’s own “to-dos” and “to-dos” one expects from others.
2. Marking things as important or outstanding among the less important items.
4. Managing deadlines and reminders, which may be associated with particular messages or other content.
5. (b only) Event-based collation of documents and discussions.

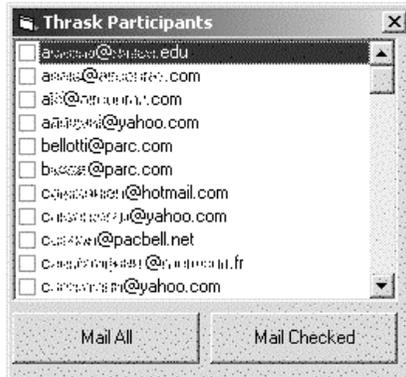
Aggregations of Information for an Overview

The fourth principle in TaskMaster is that thrasks afford an abstraction mechanism for aggregating over a collection to display useful information at the top level (addressing problem 7, shown earlier). We implemented three aggregations:

- Warning bars—These represent the nearest upcoming deadline and reminder for a given thrask (shown on the right-hand side of the top pane in Figure 6). Each bar represents 2 weeks and a proportion of the bar from the left end displayed in green shows what proportion of that time period is left before the deadline or reminder arrives. The rest of the bar is red, so the red part grows larger and the green shrinks as the date approaches (in Figure 6, the red shows up as a darker shade of grey than does the green).
- Action clusters—These represent clusters of actions associated with a thrask and are shown as miniature red or blue balls on the left-hand side of the top pane in Figure 6, next to the thrask to which they relate. Adding a new action for oneself or someone else (creating a red or blue action ball) to an item in the thrask causes a new tiny ball of the corresponding color to appear in the action cluster area.
- Task-Specific contact lists—These are task-centric lists of the names and e-mail addresses of all senders and recipients associated with items in a thrask (see Figure 7). The user can mail all the people or select multiple individuals by interacting with a pop-up dialog box containing the list attached to each thrask. The user accesses the list by clicking to the right of the thrask name.

We designed these aggregation mechanisms to give users an overview of what is important in their e-mail. So they get, at a glance, a sense of their obligations and upcoming deadlines and are able to quickly create messages to relevant collaborators without spending time searching through thrasks and inspecting individual items. Aggregations primarily deal with our final critical problem:

Figure 7. A task-specific contact list, showing all of the participants in a selected thrask, with the option of mailing some or all of them. Most addresses have been obscured to protect addressees' privacy.



7. Getting a task-oriented overview, at a glance, rather than scrolling around inspecting folders.

It is important to note that TaskMaster is a proof-of-concept prototype. We did not set out to solve all problems for e-mail and we believe that many refinements to the design are not only possible but desirable. For example, we did not solve the problem of archiving threads and removing them from view once they become inactive. Rather, the point of our prototyping efforts was to devise a task-centric view of e-mail with particular consideration to managing complex threads with incoming and outgoing obligations needing to be tracked. Our aim was to see if this repositioning of e-mail as more of a task management application had merit.

7.3. Implementation of TaskMaster

TaskMaster is implemented in Visual Basic[®] as an add-on to an Outlook client configured for the Exchange[®] mail server. It duplicates incoming messages and passes outgoing messages back to Outlook, so all traffic still appears in the user's Outlook client in the inboxes and outboxes. This implementation strategy was chosen specifically because we wanted to provide users with a "security blanket" backup of all e-mail in their Outlook client. Previously, we had found it hard to get people to switch e-mail clients, which are mission-critical tools with a great deal of legacy content and structure. The duplication approach avoided having users fear that transferring to our tool would cause them to lose all their archived e-mail when our study ended. It also

meant that should TaskMaster fail to provide any essential advanced features, users would be able to retreat back to Outlook for those features and exploit them while working with duplicates of their messages.

The Development Approach

To keep the design tied to the output of our field studies, we adapted the eXtreme Programming (XP) approach (Beck, 2000; Bellotti, Ducheneaut, Howard, Neuwirth, et al., 2002) to our needs. The fieldworkers on our team took on the role of customer and prepared stories to represent the requirements for the system, based on the seven problems listed earlier. The engineers worked in tight iterations to deliver a constantly-working prototype. We used XP's discipline to budget our resources and traded off aesthetics in favor of robust code and essential features. We also used the tool for months ourselves to manage our own e-mail. This helped us assess how well our ideas worked in practice and to refine them based on experience (and to get rid of bad ones). We also ironed out bugs and any usability problems that could be fixed with only a reasonable amount of effort.

8. EVALUATING E-MAIL AS TASK MANAGEMENT

In the design of TaskMaster, our main concern was to test the hypothesis that one can treat e-mail more seriously as task management. In particular, the features we wanted to test could not show their value in a laboratory experiment. So we wanted our study participants to use the prototype for 2 weeks in their real e-mail work. Only six participants from our in-depth, four-phase study (see Figure 8) were able to run the prototype (which depended on a particular configuration of Microsoft Office® and Outlook). So three more volunteers, BD2, R4, and R5, were recruited to make up the numbers (see Figure 8).

The TaskMaster prototype evaluation involved three steps:

1. Training on all the features of the prototype and how to use it as a task management tool, and giving participants a documentation sheet with guidelines and tips to remind them how to use it.
2. Testing participants a couple of days later to make sure they understood how to use all the features and instructing them to use the tool to manage all of their e-mail work.
3. A final brief video record and transcription of the use of and unstructured discussion of the tool, followed by a structured debriefing interview in which we asked the participants to assess each of the features

Figure 8. Participation in the evaluation of TaskMaster, showing which individuals continued on from the in-depth study of email practices.

Role	Study Phase				Prototype Use
	1	2	3	4	
Attorney (A)	—	X	X	—	—
Legal Secretary (LS)	—	X	X	—	—
Business Development (BD1)	—	X	X	—	2 days
Business Development (BD2)	—	—	—	—	1 week
Public Relations (PR)	X	X	X	X	1 week
Group Manager 1 (M1)	X	X	X	X	—
Group Manager 2 (M2)	X	X	X	X	—
Group Manager 3 (M3)	—	X	X	—	> 4 months
Patent Process Manager (PPM)	X	X	X	X	—
Researcher 1 (R1)	X	X	X	X	1 day
Researcher 2 (R2)	X	X	X	X	2 weeks
Researcher 3 (R3)	X	X	X	X	> 4 months
Researcher 4 (R4)	—	—	—	—	> 4 months
Researcher 5 (R5)	—	—	—	—	2 weeks

and rate how well the tool improved their ability to manage tasks in e-mail using Likert-type scales and freeform explanations.

Participants were given the option of terminating the study at any time if they found the tool was hindering them.

9. FINDINGS

9.1. Extent of TaskMaster Use

The plan for our evaluation was for participants to use TaskMaster for 2 weeks. Figure 8 shows the actual periods of use. Seven managed at least 1 full week. R2 and R5 used it for the full 2 weeks of the study and then returned to Outlook. M3, R3, and R4 were still using TaskMaster 4 months later, in spite of the fact that no technical support or encouragement was provided at any time after the end of the study. However, four participants did not complete the full 2 weeks and their reasons for ceasing use of the prototype early were as follows:

- R1 had to install a new version of Microsoft Office for a research project one day after starting to use TaskMaster. The new configuration was incompatible with TaskMaster, which no longer worked on her machine.
- BD1 experienced so many crashes that she stopped using TaskMaster after 2 days.
- PR experienced some addressing problems, which seemed to be an interaction between her style of use and poor feedback when real names instead of addresses were typed into “To:” fields. She reluctantly stopped using the tool after about a week, although she liked many of its features.
- BD2 did not commit to using TaskMaster exclusively (he returned to Outlook, containing his legacy e-mail, when he felt overwhelmed by work) and so unread messages in TaskMaster built up over a few days, without any user organization, making it too time-consuming for him to recover after 1 week.

It was clearly the technical limitations of TaskMaster (e.g., sensitivity to variations in PC configurations) and critical missing features (e.g., printing and feedback about address completion, mundane but developmentally-expensive features omitted due to limited engineering time resources) rather than problems with the design concepts that led to people giving up before the full 2 weeks. Indeed, TaskMaster (built in 10 person-months) has only a tiny fraction of the feature enhancements that Outlook boasts. However, even those that gave up using the prototype found the design concepts compelling; those who continued to use it, loved them. The average overall approval rating for TaskMaster was 4.11 on a Likert-type scale of 1 (*hate it*) to 5 (*love it*).

9.2. Experience of E-Mail as Task Management

All or Nothing Use

We found that inability to use TaskMaster exclusively, for whatever reason, limited its success. E-mail users invest heavily in the features of their client and the structures of folders that they create over time, relying on them to systematize and keep track of work. Using two tools simultaneously that accumulate the same messages (as TaskMaster and Outlook were doing) requires keeping both sets of e-mail up to date; double the investment. BD1, who gave up using the tool after 2 days (due to instability of the prototype), summed up this point most clearly:

I continued to use Outlook for global addressing contacts, calendar those sort of things. Not having these things really made writing e-mail cumbersome. [...] I

would give it a higher rating, but I had to go back and folder everything in Outlook. [...] I felt I had to do the work twice.

Similarly, PR experienced severe addressing problems and BD2 also reported that a lack of key features (printing and formatting) caused him to keep going back to Outlook. For PR and BD2, who both used the tool for a week, it became difficult to continue with TaskMaster not because the concept of e-mail as task management failed but because TaskMaster lacked standard features that they wanted and because switching back and forth to Outlook to use those features was impractical.

The five who were able to get by without certain features, and who did not experience serious technical problems, were all able to complete the study. From the outset, we had modest expectations about the chances of getting people to use TaskMaster for the full 2 weeks, given its lack of features compared to what our users were used to with Outlook. So we were delighted when we discovered that three of the five participants who completed the study did not want to switch back to Outlook despite TaskMaster's limitations. Clearly, even with the handicap of a very limited feature set, TaskMaster had some important advantages over Outlook. In the following sections, we take each of our design principles in turn and discuss how its embodiment in our prototype was received by our users and how its implementation might be refined.

Threaded Task-Centric Collections

TaskMaster's thrasks were both reliable at collecting threaded messages, attachments, and links and a successful means of organizing message content. As expected, preserving the context of messages was useful (Cadiz et al., 2001; Eklundh & MacDonald, 1994; McDaniel et al., 1996): "The high level good thing is that I think that it is surprisingly useful to me that when something arrives I actually see it in the context of the conversation" (R2).

Our users rated both organizing and deleting content within the thrask model with an average 4.2 on a scale of 1 (*hate it*) to 5 (*love it*). As BD1 put it, "TM does a generally good job of thrasking, so I've only had to merge it a few times, so it's great." Users combined thrasks more often than they broke them apart, suggesting that topic drift was perhaps not as common as we might have anticipated in light of our previous fieldwork (Ducheneaut & Bellotti, 2001a). However, people still found splitting thrasks useful, even if they did it infrequently.

Providing the ability to intervene in what would otherwise be a simple message threading function clearly confers an advantage over simple threading. Combining documents and links added to that advantage:

It's just nice to be able to have the control over mixing [...] related things together, even though they might not be sort of the identical kind of thing. And I've also taken a bunch of things out of my Outlook [...] that would be applicable to a particular thrask. (M3)

Users' Proposed Design Refinements. Some TaskMaster users suggested that being able to archive thrasks would have been an advantage for the long term. One user stated that being able to create subthrasks would be useful too. It seems that, ideally, rather than creating complex hierarchies of subthrasks, they could be blended with hierarchical folders for archival purposes.

Another refinement, which was more a matter of a missing feature, was to have the sender of the most recent message in a thrask be visible in TaskMaster's top pane. We had previously established in interviews that the sender is the critical determinant to the recipient of whether a message is likely to be important and should be made immediately visible.

Equality for All Content

The thrask model was made considerably more powerful by elevating documents and links to be first-class citizens and displaying them in the preview pane when selected in the item list view. In TaskMaster, important documents and links are much easier to get at when they are needed:

Last week I had this meeting that had, I don't know, about 4 different responses and about 6 different documents and it was really helpful to be able to have them all in one place. And to be able, during the conference call, to bring up everybody's message [...] and have the attachments [...] it's just really easy to click and not have to open this message and then go down and click here and wait for the application to open.... (M3)

TaskMaster also saves time that is normally wasted opening uninteresting attachments to see what they are:

I liked that a lot because the difference here [In Outlook] is for example, this is an image but... this image is tiny; it's not very important. Now I'm clicking and I'm clicking and it's just... there's so many steps, for just this [wasting time opening an unimportant image] and I could have seen that in my e-mail [in TaskMaster]. (PR)

Content equality in listing and previewing was rated as a 4 on our Likert-type scale. However, there was clearly a need to refine it (discussion follows).

Users' Proposed Design Refinements. The biggest problem with the content management was that the layout of the three panes of TaskMaster was fixed and the bottom pane, in particular, was not large enough. It was clear that users would prefer to be able to pick a personal window configuration. Double-clicking on items does open them in their own separate window, which can be sized as desired, but this takes us back to wasting time on window management. Users needed a better layout solution for all three of the TaskMaster view panes.

Another proposed refinement was to provide a fully functional application preview pane so that items could be manipulated without ever having to open an application.

Task-Centric Meta-Information for All Items

Although users liked most of the task-centric meta-information such as deadlines, reminders, action balls, color-coding, and iconic flags, these items were not quite as successful as we had hoped (rated between 3 and 4 on our Likert-type scale). The deadlines, reminders, and action balls were the most successful, being used occasionally and rated around 4:

When I have a lot of thrasks like this that are mainly just reminders, it's helpful for me to quickly look and go oh I'm the one that needs to do something here. [...] it is also very handy for me to keep track of, OK this is here because I'm waiting for someone else to give me the information. (M3)

There is a [...] bizdev [thread] where somebody asked a question about market size, and in almost all circumstances I would have forgotten ever to reply to that, but it's still flagged, it's overdue, I put a little red dot also, it means I owe something to somebody. (BD2)

Users' Proposed Design Refinements. The main problem with our meta-information was that, although users liked the ideas, they found that better integration with information management resources would have helped. For example

Can you extract all the messages or the thrasks that are color-coded? [...] It would reiterate this kind of color-coding with the bars. (R5)

One user proposed being able to attach meta-information such as actions to entire thrasks and then to provide more details about that information at the item level. Another proposed an idea that we did have but never implemented (see, however, Rohall & Gruen, 2002):

See if I had something here where I could write a note to myself, what I wanted to do, not just put a reminder but write a note to myself. (R4)

We concluded that we were heading in the right direction with these features but many more enhancements were needed to make meta-information more effective for users. We believe calendar integration, searching by meta-information, and additional aggregations will improve the value of meta-information.

Aggregations of Information for an Overview

The aggregations made meta-information much more useful by providing an overview in the top pane. The deadline and reminder aggregations (warning bars) were especially popular (our most popular feature, rated at an average of 4.4 on our scale).

I liked the visualization bars a lot, that's something for which there is no counterpart in Outlook. Seeing the growing red, just having an idea at a glance. (R1)

They're a visual reminder [...] I could see sort of the slack time that I have, the relation of the green to red gave me an indication of how urgent it was. (R3)

This is visually so there I mean, green or red!!! I just thought it was terrific to walk in, in the morning and to see the change over time and be able to know what was due Monday morning, versus Tuesday morning and to be able to see that transition and go yes I did, yes I did, yes I did, no I didn't [pointing at the bars in P1]. (PR)

In most thrasks, we saw only one action ball, so the aggregation simply served as a reminder that there was an action associated with a thrask. It may be that the number of actions associated with a thrask is not important as an aggregation. Simply knowing there is some action could be enough, because the user only needs to have his or her attention drawn to the thrask to open it and be reminded of any actions by the message headers and action balls inside.

The task-specific contact lists (pop-up lists of all those involved in any messages in a thrask; rated on average at 3.5 on our scale) were severely handicapped due to implementation problems; quirks in Microsoft's e-mail addressing scheme often led to inclusion of multiple strange addresses for each contact. Some users loved this feature, but others did not:

When I saw it I knew it was the right thing but it has the usability thing about the addresses I can't read. And the other thing is sometimes the list has duplicates. (R2)

It would have taken too much work to resolve this problem in time for our study but, despite negative reactions to the imperfect implementation, the idea itself seems very promising based on encouraging feedback from our users.

Our users showed us that aggregations are useful if they are of the right kind. For example, warning bars gave people the ability to see at a glance which thrask contained the most urgent item. This helped in prioritizing e-mail work. However, the action balls did not convey which actions were most important, so although they gave a sense of “to-do-ness,” they did not help much in planning work. Perhaps R4’s suggestion of adding notes might have improved “to-do” aggregation. One could imagine users adding a couple of words to actionable items, which could be propagated up to the thrask level, like the warning bars.

Some users also pointed out that our task-centric address lists lacked the essential property of being able to distinguish between “To:,” “Cc:,” and “Bcc:” fields, which made them much less useful than they could have been.

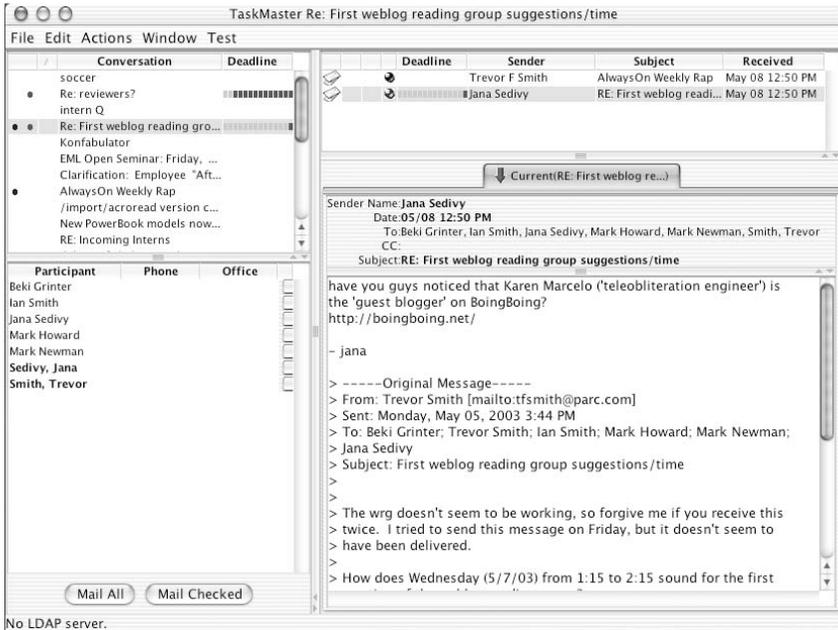
Summary of Findings

Overall, our evaluation of TaskMaster showed that positioning e-mail as task management is something that users do find compelling. Our prototype suffered from many technical implementation problems and a limited feature set. In spite of this, the fieldwork-driven design principles—“threaded task-centric collections,” “equality for all content,” “task-centric meta-information for all items,” and “aggregations of information for an overview”—were compelling enough to outweigh the limitations for some users who still prefer to use the prototype instead of Outlook. Further, all our users’ comments have contributed to a number of ideas for how our design principles can be refined in future implementations; for example, we have worked hard to improve the contacts feature which was seen as a poor implementation of a good idea in the original prototype.

10. ONGOING WORK AND CONCLUSIONS

Our evaluation of TaskMaster showed that it is a robust solution to the problem of dealing with complex tasks in e-mail. Based on the success with, and user feedback on, the first prototype, another version of the tool is now being implemented in Java as a stand-alone client (see Figure 9). This version, presently incomplete, has an improved layout scheme and a more aesthetic appearance, but still preserves the same fundamental ideas, although thrasks have been renamed more appealingly as conversations. In particular, we have augmented our aggregate information visualizations, discussed in sec-

Figure 9. The new TaskMaster client with improved layout, resizable panes, and permanently visible contact lists.



tions 7.2 and 9.2. We have added warning bars to individual items (top-right pane in Figure 9) as well as maintaining the aggregate (nearest deadline) warning bars at the top level (top-left pane in Figure 9). We are also working on making contacts a permanently visible resource updated from our organization's server-based global address book. Finally, we are currently investigating ways to improve on the action items by providing a variety of richer resources for visualizing tasks and actions associated with e-mail.

Our research shows that a significant component of e-mail overload is the quality of messages rather than simply the quantity. Messages involved in complex extended, and especially interdependent, tasks involving coordination with others contribute heavily to a sense of overload and are difficult to manage with current e-mail clients. Tools that simply help to classify messages for archiving purposes do not solve the problems associated with these kinds of tasks. We believe that it is possible to significantly and positively affect e-mail users' experience by embedding task management resources directly in the inbox, where they are most needed, as well as breaking down the barriers between the various components of contemporary e-mail applications. The small set of features we have built into and tested in our TaskMaster pro-

prototype appears to be a strong foundation for a radical (and long overdue) overhaul of e-mail's user interface. It is also a clear indication that life in the e-mail habitat should be rethought not in terms of messaging but rather in terms of the various activities users are trying to accomplish through that activity. As we noted in our users' comments, however, there is much work left to do to perfect this vision. Therefore, we are currently working on pushing some of our concepts further in a new version of the prototype.

NOTES

Acknowledgments. We dedicate this article to the memory of our beloved colleague Mark Howard who passed away in England in 2004. We also gratefully acknowledge the generosity of our study participants who gave us so much time and information, despite, in many cases, being very "overloaded."

Authors' Present Addresses. Victoria Bellotti and Nicolas Ducheneaut, Palo Alto Research Center, 3333 Coyote Hill Road, Palo Alto, CA 94304. E-mail: bellotti@parc.com, nicolas@parc.com. Ian Smith, Intel Research Seattle, 1100 NE 45th Street, 6th Floor, Seattle, WA 98105. E-mail: ian.e.smith@intel.com. Rebecca E. Grinter, GVU Center — TSRB College of Computing, Georgia Institute of Technology, 85 Fifth Street, NW Atlanta, GA 30332-0760 USA. E-mail: beki@cc.gatech.edu.

HCI Editorial Record. First manuscript received June 2, 2003. Revision received March 14, 2004. Accepted by Paul Moody. Final manuscript received October 26, 2004. —*Editor*

REFERENCES

- Bälter, O. (2000). Keystroke level analysis of email message organization. *Proceedings of the CHI 2000 Conference on Human Factors in Computing Systems*. New York: ACM.
- Bälter, O., & Sidner, C. (2000). *Bifrost Inbox Organizer: Giving users control over the inbox* (Technical Report 00-08). Yorktown Heights, NY: IBM Watson Research Center.
- Beck, K. (2000). *Extreme programming explained*. Reading, MA: Addison-Wesley.
- Bellotti, V., Dalal, B., Good, N., Flynn, P., Bobrow, D., & Ducheneaut, N. (2004). What a to-do: Studies of task management towards the design of a personal task list manager. *Proceedings of the CHI 2004 Conference on Human Factors in Computing Systems*. New York: ACM.
- Bellotti, V., Ducheneaut, N., Howard, M., Neuwirth, C. M., Smith, I., & Smith, T. (2002). Innovation in extremis: Evolving an application for the critical work of email and information management. *Proceedings of the DIS 2002 Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*. New York: ACM.
- Bellotti, V., Ducheneaut, N., Howard, M., & Smith, I. (2002). FLANNEL: Adding computation to electronic mail during transmission. *Proceedings of the UIST 2002 Conference on User Interface Software and Technology*. New York: ACM.

- Bellotti, V., Ducheneaut, N., Howard, M., & Smith, I. (2003). Taking email to task: The design and evaluation of a task management centered email tool. *Proceedings of the CHI 2003 Conference on Human Factors in Computing Systems*. New York: ACM
- Bellotti, V., & Smith, I. (2000). Informing the design of an information management system with iterative fieldwork. *Proceedings of the DIS 2000 Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*. New York: ACM.
- Boone, G. (1998). Concept features in Re:Agent, an intelligent email agent. *Proceedings of the Agents 98 Second International Conference on Autonomous Agents*. New York: ACM.
- Cadiz, J. J., Dabbish, L., Gupta, A., & Venolia, G. D. (2001). *Supporting email workflow* (Technical Report MSR-TR-2001-88). Redmond, WA: Microsoft Research.
- Czerwinski, M., Smith, G., Regan, T., Meyers, B., Robertson, G., & Starkweather, G. (2003). Toward characterizing the productivity benefits of very large displays. *Proceedings of the INTERACT 03 Conference on Human-Computer Interaction*. Amsterdam: IOS Press.
- DeSanctis, G., & Poole, M. S. (1994). Capturing the complexity in advanced technology use: Adaptive structuration theory. *Organization Science*, 5, 121-147.
- Denning, P. (1982). Electronic Junk. *Communication of the ACM*, 25(3), 163-165.
- Ducheneaut, N., & Bellotti, V. (2001a). Email as habitat: An exploration of embedded personal information management. *Interactions*, 8, 30-38.
- Ducheneaut, N., & Bellotti, V. (2001b). *A study of email work activities in three organizations*. Unpublished manuscript, Palo Alto, California, Research Center.
- Ducheneaut, N., & Bellotti, V. (2003). Ceci n'est pas un objet? Talking about objects in email. *Human-Computer Interaction*, 18, 85-110.
- Eklundh, K. & Macdonald, C. (1994). The use of quoting to preserve contents in electronic mail dialogues. *IEEE transactions on professional communication*, 37(4), 197-202.
- Fisher D., & Moody, P. (2001). *Studies of automated collection of email records* (Technical Report ISR-02-04). Irvine: University of California Irvine, Institute for Software Research.
- Gartner. (2001). *How to rescue e-mail systems from occupational spammers* (Archive Note FT-13-5977). Stamford, CT: Author.
- Hollan, J., & Stornetta, S. (1992). Beyond being there. *Proceedings of the CHI 92 Conference on Human Factors in Computing Systems*. New York: ACM.
- McDaniel, S. E., Olson, G. M., & Mage, J. C. (1996). Identifying and analyzing multiple threads in computer-mediated and face-to-face conversations. *Proceedings of the 1996 Conference on Computer Supported Cooperative Work*. New York: ACM.
- Mackay, W. E. (1988). More than just a communication system: Diversity in the use of electronic mail. *Proceedings of the CSCW 88 Conference on Computer-Supported Cooperative Work*. New York: ACM.
- Mintzberg, H. (1973). *The nature of managerial work*. New York: Harper & Row.
- Mock, K. (2001). An experimental framework for email categorization and management. *Proceedings of SIGIR 2001 24th Annual International Conference on Research and Development in Information Retrieval*. New York: ACM.
- Reder, S., & Schwab, R. (1990). The temporal structure of cooperative activities. *Proceedings of the CSCW 90 Conference on Computer Supported Cooperative Work*. New York: ACM.

- Rohall, S. L., & Gruen, D. (2002). ReMail: A reinvented email prototype. *Demo at the CSCW 02 Conference on Computer Supported Cooperative Work*. New York: ACM.
- Segal, R. B., & Kephart, J. O. (1999). MailCat: An intelligent assistant for organizing email. *Proceedings of the Agents 99 Third Annual Conference on Autonomous Agents*. New York: ACM.
- Sproull, L. (1984). The nature of managerial attention. In L. S. Sproull & P. D. Larkey (Eds.), *Advances in information processing in organizations* (pp. 9–27). Greenwich, CT: JAI.
- Sproull, L., & Kiesler, S. (1991). *Connections: New ways of working in the networked organization*. Cambridge, MA: MIT Press.
- Takkinen, J., & Shahmehri, N. (1998). CAFE: A conceptual model for managing information in electronic mail. *Proceedings of the HICSS-31 Thirty-First Hawaii International Conference on System Sciences*. Washington, DC: IEEE Computer Society.
- Takkinen, J., & Shahmehri, N. (1999). Task-oriented restructuring of an application domain: A multi-agent architecture for doing things in Internet email. *Proceedings of the HICSS-32 Thirty-Second Hawaii International Conference on System Science..* Washington, DC: IEEE Computer Society.
- Venolia, G., Dabbish, L., Cadiz, J., & Gupta, A. (2001). *Supporting email workflow* (Technical Report MSR-TR-2001-88). Redmond, WA: Microsoft Research.
- Whittaker, S., & Sidner, C. (1996). Email overload: Exploring personal information management of email. *Proceedings of the CHI 96 Conference on Human Factors in Computing Systems*. New York: ACM.