Mental Workload in Personal Information Management: Understanding PIM Practices Across Multiple Devices

Manas Tungare

A Dissertation presented to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Computer Science and Applications

Committee:

Manuel A. Pérez-Quiñones, Chair
Stephen H. Edwards
Edward A. Fox
Steven R. Harrison
Tonya L. Smith-Jackson

March 25, 2009; Blacksburg, VA, USA.

Keywords: Personal Information Management, Multiple Devices, Mental Workload.

Mental Workload in Personal Information Management: Understanding PIM Practices Across Multiple Devices

Manas Tungare

Abstract

Multiple devices such as desktops, laptops, and cell phones are often used to manage users' personal information, such as files, calendars, contacts, emails, and bookmarks. This dissertation presents the results of two studies that examined users' mental workload in this context, especially when transitioning tasks from one device to another. In a survey of 220 knowledge workers, users reported high frustration with current devices' support for task migration, e.g. accessing files from multiple machines. To investigate further, I conducted a controlled experiment with 18 participants. While they performed PIM tasks, I measured their mental workload using subjective measures and physiological measures. Some systems provide support for transitioning users' work between devices, or for using multiple devices together; I explored the impact of such support on mental workload and task performance. Participants performed three tasks (Files, Calendar, Contacts) with two treatment conditions each (lower and higher support for migrating tasks between devices.)

I discuss the following findings in this dissertation: workload measures obtained using the subjective NASA TLX scale were able to discriminate between tasks, but not between the two conditions in each task. Task-Evoked Pupillary Response, a continuous measure, was sensitive to changes within each task. For the Files task, a significant increase in workload was noted in the steps before and after task migration. Participants entered events faster into paper calendars than into an electronic calendar, though there was no observable difference in workload. For the Contacts task, task performance was equal, but mental workload was higher when no synchronization support was available between their cell phone and their laptop.

Little to no correlation was observed between task performance and both workload measures, except in isolated instances. This suggests that neither task performance metrics nor workload assessments alone offer a complete picture of device usability in multi-device personal information ecosystems. Traditional usability metrics that focus on efficiency and effectiveness are necessary, but not sufficient, to evaluate such designs. Given participants' varying subjective perceptions of these systems and differences in task-evoked pupillary response, aspects of hot cognition such as emotion, pleasure, and likability show promise as important parameters in system evaluation.

All text, illustrations, graphs, tables, figures, photos and other supplementary material included in this dissertation were created and typeset for publication by the author in Adobe Caslon Pro and Myriad Pro font faces using the free LaTeX document preparation system. Statistical analyses and graphs were obtained with scripts written for the R software environment for statistical computation. Additional illustrations were created in OmniGraffle and iWork '09 on Apple Mac OS X 10.5 Leopard.

To my parents and grandparents ...

आजोबा-आजी, काका आजोबा-आजी, व डॅडी-ममा, यांना ...

Acknowledgments

A dissertation is neither a destination unto itself, nor is it the journey of a solitary traveler. There are always those who meet him along the way, offer him their company, show him the light in times of uncertainty, dance in times of joy, and above all, celebrate the journey with him. I have been blessed to have met several extraordinary people who have influenced me throughout this expedition.

My dissertation committee brings together insights and experiences from a wide variety of backgrounds. In discussions I have had with them, there have been several *'Aha!'* moments, triggered by the synthesis of their perspectives on several ideas. My primary advisor, Dr. Manuel Pérez-Quiñones, sparked my interest in multi-device interfaces and personal information management. From our early collaboration through independent studies, classes, and research, he has helped me navigate the breadths and depths of topics I was interested in, to pick one with lasting impact and valuable long-term contributions. Dr. Tonya Smith-Jackson engaged me in the exploration of mental workload, a topic I had been introduced to in her class a few years earlier. With Prof. Steve Harrison, I had fruitful discussions about the role of places and context in information ecosystems. Dr. Ed Fox gave excellent feedback on the choice of tasks in the experiment, and Dr. Steve Edwards pointed out aspects of the statistical analyses that could potentially be problematic. I am indebted to all of them for guiding this dissertation to its conclusion.

I have enjoyed several conversations about research, personal information management practices, and lots more with my friend and colleague, Pardha Pyla. Some of the ideas that I have explored in detail in this dissertation came from our early discussions, and in the years since, he has been one-third of the Friday afternoon chats over pizza with me and Dr. Pérez. While formulating my statistical analyses, I frequently turned to my friend, Ranjana Mehta, to make sure that my work was not only statistically correct, but also would lead to valuable insights into the data.

PIM has been a frequent topic of conversation among us members of the PIM Lab: Ben, Ricardo, Sameer, Pardha, and I have shared several personal stories and research discussions in the lab. During my annual west-bound summer migrations to Google, I worked on real-world problems which grounded this work in the practice of HCI. While working with Dr. Bill Schilit, I realized the efficiency of the release-early/release-often approach, which I applied to my experiment design, refining it during several passes of pilot studies. Over the past few years, I have had the opportunity to discuss my research with others in the Personal Information Management community, thanks to the annual PIM workshops. Discussions with Rob Capra (also an alumnus of my advisor, and an office neighbor in my second year), Deborah Barreau, William Jones, Jamie Teevan, Rick Boardman, and Mary Czerwinksi have all influenced this work.

The Center for HCI at Virginia Tech provided the eye tracking equipment for this experiment, which was crucial to the pupillometric measurements I performed. Special thanks to Chreston Miller for familiarizing me with the eye tracker. The CHCI, Dept. of Computer Science's Graduate Travel Fund, Virginia Tech Graduate Student Assembly's Travel Fund Program, and the National Science Foundation funded portions of many of my trips to attend conferences and enabled me to interact with others in the field. The excellent administration of the Computer Science department has always been helpful in whatever I asked of them: a big thank you to Dr. Naren, Dr. Ribbens, Tess, Carol, Rachel, Melanie, Jessie, Ginger, Julie, Gen, Jody, and Lucy for shielding us from the bureaucracy at upper levels of the university. On a sombre note, Dr. Kim Beisecker, Director of Cranwell International Center, deserves my gratitude for her extraordinary strength on the night of April 16, 2007 that we spent at Virginia Tech Inn awaiting identification of our friends who perished through no fault of theirs.

Over the last four years, many friends have made the journey truly enjoyable: the 'Friday Night Bunch' helped break the monotony of the week, and welcomed weekends as they truly should be. Pardha, Hari, Uma, Bhawani, Tejinder, Mara, Claudio, Sarah, Laurian, Edgardo, Rhonda, Jason, Shahtab, Yonca, Ergun, Stacy, Sameer, Meg, Wes — the list goes on. My house-mates and friends (and their respective visiting spouses) have been great company for late night philosophical chats about nothing in particular: Vivek, Mansi, Amar, Aarati, Siya, Brijesh, Parag, Amit, Rachana, Ashish, Sunil, Shrirang, Deepti, Harsh & Neha. Neeraj not only has been a close friend, but also a gracious host in Washington DC during my frequent weekend getaways. Those who know me are aware of my wanderlust; my travel companions made sure I took a break every few months to recharge my batteries. Many of them are my oldest and closest friends, from grade school, high school and college: Kavita, Hemali, Mihir, Sharvari, Laukik, Amit, Supriya, Rohan, Alok, Kashmira, Ajay, Vinita, Aparna.

During the last four years and a half, I have had an extremely enriching experience working with my advisor, Dr. Pérez. He encouraged me to come up with my own ideas, shaping them along the way, instead of handing down a project specification. It is this freedom to explore that enabled me not only to get a degree, but also to learn all the valuable lessons along the way that make the degree worth it. Like the old adage, he did not hand me a fish, but instead taught me how to fish. Together, we've worked on publications not only at the office, but also over instant messaging past midnight, on baseball fields, on road trips to conferences, and via Facebook and Twitter. I will treasure our collaboration through the rest of my professional career, and hope to continue working with him well past this milestone.

Finally, this milestone is due in no small part to the support and encouragement of my parents and grandparents: they have always taught me to pursue my dreams and ambitions, and they are likely more proud of this achievement than I am. My parents have been the bedrock upon which my dreams have been built. They ignited in me a love for science at an early age, and introduced me to computers when I was around 10 – still a veritable novelty in 1990. Ajoba (my grandfather) has seen me progress through this entire journey, from taking me to the Science Center as a kid, to seeing me off as I proceeded to the airport. Aaji (my grandmother) gave me twice the love, since my other Aaji already had left this world before I came into being. But, during the last few years, we lost her, though I'm sure she is happy with Kaka Ajoba wherever they both are.

To all of you who have touched my life in various ways, thank you!

Contents

List of Figures

List of Tables

Chapter 1

Introduction

> "Now that we've built computers, first we made them room-size, then desk-size and in briefcases and in pockets, soon they'll be as plentiful as dust — you can sprinkle computers all over the place. Gradually, the whole environment will become something far more responsive and smart, and we'll be living in a way that's very hard for people living on the planet just now to understand."
>
> — *Douglas Adams.*
> *Posthumously published in [Adams, 2002]*

Computers did become smaller and plentiful over the years, but our interaction with them hardly merits a comparison to 'sprinkling'. If anything, they have contributed to increased stress and frustration when technology falls short of a user's expectations and intentions.

Information is being disseminated much faster than we can assimilate it. Our tools are not adapting fast enough to keep pace with the need for ubiquitous access to information. A large sector of the economy is devoted to managing information, and information overload threatens our effectiveness. Even at home, we are inundated with information as we manage an ever-increasing library of documents, to-do lists, digital music, digital photos, and others. All of this causes stress and increases mental workload as we struggle to stay in control of our information.

One of the biggest challenges of our time is to control effectively the management of personal information. We have developed amazing capabilities to record, store, and transmit massive quantities of information with minimal effort; however this has relegated us to file clerks [Dumais and Gates, 2003] and part-time librarians of our own personal information. In spite of the ease of recording, creating, receiving, storing, and accumulating digital materials, it is difficult to manage and use them sensibly [Gemmell et al., 2002, Czerwinski et al., 2006]. With time, the amount of information generated by humans can only increase, while human attentional resources have remained constant [Levy, 2005].

At the same time, advances in computer hardware have led to the miniaturization of technology that places several portable information devices at our disposal. It is common for a lot of people to carry a laptop computer or a cell phone as they go about their everyday business, outside the usual contexts of an office or a home [Dearman and Pierce, 2008, Tungare and Pérez-Quiñones, 2008b], and to expect productive work output when mobile. However, the current state-of-the-art in information management solutions sends these users into a frenzy trying to locate the most current version of their slide shows, the documents they sent around for review, and the phone number of the person they need to call right now.

When several devices are used together, as in Personal Information Ecosystems [Pérez-Quiñones et al., 2008], a user needs to focus attention on various tasks at the same time, or in quick succession. In traditional single terminal computer systems, the majority of a user's attentional and cognitive resources would be focused on the terminal while performing a specific task. However, in an environment where multiple devices require intermittent attention and present useful information at unexpected times, the user is subjected to different mental workloads.

In this dissertation, I examine the impact of multiple devices on a user's personal information management tasks. Specifically, I am interested in how different designs of multi-device personal information management systems affect the mental workload and frustration caused to users.

1.1 Problem Domain

This work is situated at the intersection of three areas of enquiry within the broader domain of human-computer interaction. I focused my investigations on tasks in the domain of **Personal Information Management** (introduced in §1.1.1; details in §2.2); this includes tasks such as file management, calendar management and contact management. These tasks were performed by users using **Multiple Devices** (introduced in §1.1.2; details in 2.3), an area in its own right that has been studied widely from the point of view of interaction, but less so from the point of view of information. To measure users' performance on these tasks, I use theories and methods developed in the field of **Mental Workload Assessment** (introduced in §1.1.3; details in 2.5).

1.1.1 Personal Information Management

Given that one's personal information exists in a continuum, and that it spans all aspects of one's life, deriving a comprehensive definition for it is challenging. Jones [Jones, 2008] covers the salient aspects of personal information by defining it as information that is controlled by or owned by us, about us, directed towards us, sent (posted, provided) by us, (already) experienced by us, or relevant (useful) to us. Various aspects of personal information management have been studied in the literature (see chapter 2). They include studies of various types of personal information, approaches and user traits (pilers versus filers [Malone, 1983], browsing versus searching [Teevan et al., 2004], etc.) and cross-project information management [Boardman et al., 2003, Bergman et al., 2006]. Individual information collections have been studied, e.g. files [Barreau and Nardi, 1995], calendars [Kelley and Chapanis, 1982, Payne, 1993, Tungare and Pérez-Quiñones, 2008a], contacts [Whittaker et al., 2002a, Whittaker et al., 2002b], etc. A problem in evaluating PIM tools or systems is that personal information is, by definition, personal [Kelly, 2006]. Thus, it is difficult, or close to impossible, to develop reference tasks that can be performed by multiple users to test multiple tools and approaches [Kelly and Teevan, 2007]. There is a pronounced lack of measurement techniques that are known to work across tasks, across tools, and across experiments [Teevan and Jones, 2008].

1.1.2 Multi-Device User Interfaces

The research discipline of multi-device user interfaces has extensively studied how applications may be written to run on many platforms [Thevenin and Coutaz, 1999, Florins and Vanderdonckt, 2004, Denis and Karsenty, 2004, Ali et al., 2005], but not much work has focused on understanding how users access or manage their information across multiple devices. Research in this area has followed a task-oriented approach rather than an information-oriented approach. The importance of following an information-oriented approach has been well-highlighted [Fidel and Pejtersen, 2004].

The impact of such multiple devices on personal information management is more than that of the individual devices alone. In a way, these devices are the analogues of various organisms that constitute a biological ecosystem [Pérez-Quiñones et al., 2008]. When used together, e.g. at a desk, these devices compete for a user's attention, and require valuable mental resources to be attended to. The influence of a multi-device environment on the user's mental workload, and how it affects operator performance under these conditions has not been studied in detail.

1.1.3 Mental Workload Assessment

In the rest of this dissertation, I will use the following general definition of Mental Workload: *Mental workload is defined as "[...] that portion of operator information processing capacity or resources that is actually required to meet system demands."* [O'Donnell and Eggemeier, 1986]. Workload can be measured in several ways: via performance-based assessment techniques (§2.5.2), via subjective workload assessment techniques (§2.5.3) or via physiological workload assessment techniques (§2.5.4).

While operator performance in a particular task situation can be measured directly by performance metrics, (e.g. the time taken to perform an experimental task, or the number and severity of errors in task performance), they cannot be used to predict performance for an unknown task [Wilson and Eggemeier, 2006]. Subjective workload assessment techniques such as NASA

Task Load Index (NASA TLX) [Hart and Staveland, 1988], the Subjective Workload Assessment Technique, [Reid et al., 1982], and Workload Profile [Tsang and Velazquez, 1996] are used to provide an estimate of mental overload. It is generally assumed that workload is related to operator performance such that low to moderate levels of workload are associated with acceptable levels of operator performance [Wilson and Eggemeier, 2006]. Mental Workload has often been studied in high-stress critical work environments, but not in office-type work environments with knowledge workers. In my dissertation, I examine the applicability of workload assessment in PIM tasks.

1.2 Motivation

As we amass vast quantities of personal information, managing it has become an increasingly complex endeavor. The emergence of multiple information devices and services such as desktops, laptops, cell phones, PDAs and cloud computing adds a level of complexity beyond simply the use of a single computer. In traditional single terminal computer systems, the majority of a user's attentional and cognitive resources are focused on the terminal while performing a specific task. However, in an environment where multiple devices require intermittent attention and present useful information at unexpected times, I hypothesize that the user is subjected to different mental workload.

In 2007, I conducted a survey study [Tungare and Pérez-Quiñones, 2008b] to understand the use of multiple devices in personal information and identify common tasks, activities, devices, patterns, device affinities, and problems in their use. This study, its analysis and findings are reported in section §3.2 of this dissertation. Many findings were close to what we expected: that users preferred laptop computers over desktops; several users owned and regularly used more than two computers, plus a cell phone, a digital camera, etc. However, a surprisingly high number of users reported chronic problems in using multiple devices together for managing their tasks. Synchronization issues between information collections on two or more machines were cited as the most common problem. Sprouting from this investigation, I decided to examine this problem deeper — whether the level of system support for such basic processes as information migration affects user performance and workload. Since several common tasks were identified by survey participants, I proceeded to explore this for three separate information collections (files, calendars, and contacts).

In the survey conducted, several users were very passionate in reporting horror stories of their use of multiple devices. Many of them had faced issues ranging from not being able to contact a person when they needed to, to complete data loss when transferring data between devices. The tone of the narration of their experiences in response to a questionnaire revealed that there was a serious undercurrent of frustration at the status quo in personal information management tools. While current usability metrics are able to provide evaluations of interfaces based on objective qualities such as efficiency and performance, other non-traditional factors such as user enjoyment, acceptance, happiness and satisfaction are not measured or reported in studies.

The traditional definition of usability, according to the International Standards Organization

[International Standards Organization, 2008] describes it as the *"the effectiveness, efficiency and satisfaction with which specified users can achieve specific goals in particular environments."* Dillon argues [Dillon, 2002a] that traditional usability metrics measure efficiency, but that may not correspond well with users' goals in using a particular system. Specifically, he comments that usability is necessary but not sufficient to ensure good design [Dillon, 2002b]. He proposes the extension of the ISO approach to usability to include components such as user satisfaction and elements of affect.

Traditional usability metrics focus on the user interaction with a single device, or with multiple devices independently. What happens at the transition between the two devices is not only difficult to measure, but also hard to quantify. This is especially troubling because most of the problems reported by users stemmed from their inability to migrate tasks successfully across devices, but rarely in their lack of ability to use a single device effectively. It appeared that the source of the problems and frustration was rooted in transitional states that required planning and coordination on part of the user, and lack of support from the system. Other factors such as product- or brand-preference have also been shown to impact usability ratings [Park et al., 2006].

In the next few sections, I describe the specific research questions I sought to answer, the approach I took, and the contributions I expect from this work.

1.3 Research Questions & Approach

The principal issue I was interested in studying was whether PIM tasks that typically are performed using multiple devices together result in high workload. Do they lead to an increased perception of task difficulty and/or cause users to switch to workarounds that result in lower workload? This section describes the details of the three specific, measurable research questions I set out to answer.

1.3.1 RQ 1: Mental Workload across Tasks and Levels of Support

Research Question

What is the impact of (1) different tasks and (2) different levels of system support for migrating information, on the workload imposed on a user?

Certain tasks require more attentional resources than others, and may result in increased mental workload, while certain other tasks may be straightforward and may require fewer mental resources. What is the variability in the subjective assessment of mental workload for these tasks? Systems provide varying levels of support for moving information mid-task from one device to another. What is the effect on workload of the level of system support for such migration?

Systems differ in the level of support they provide for pausing a task on one device, and resuming it on another [Pyla et al., 2006]. A goal of my research is to examine if mental workload at the point of transition is correlated with the level of system support available for the sub-task of transitioning. Miyata and Norman hypothesized [Miyata and Norman, 1986] and Iqbal et al. [Iqbal and Bailey, 2005] demonstrated that within a single task, mental workload decreases at sub-task boundaries. But when a sub-task is performed on a different device than the first, what are the changes in mental workload? Is it possible to reduce mental workload in a task by supporting task migration better?

Hypothesis

I hypothesize that the variability in workload imposed by dissimilar tasks will be high. The level of support provided by the system for task migration affects mental workload: higher level of support would lead to lower levels of workload and vice-versa.

In addition, I hypothesize that at sub-task boundaries where transitions occur between devices, mental workload rises just before the transition and returns to its normal level a short duration after the transition is complete.

Approach

To verify this hypothesis, I conducted an experiment to measure mental workload for three different tasks, related to Files, Calendar and Contacts, at two levels of system support for information migration. After each task, I requested participants to complete a subjective workload evaluation using the NASA TLX workload assessment technique (§2.5.3). During each task, participants wore an eye tracker which measured their pupil radius, which was used as a continuous estimate of mental workload (§2.5.4). Results are presented in section §3.9.

1.3.2 RQ 2: Operator Performance at Different Levels of System Support

Research Question

How is user performance impacted at differing levels of system support for performing tasks across multiple devices? To evaluate this, I simulated two conditions for each task; in each case, the L0 condition offered a lower level of support for migrating tasks between devices than the L1 condition. How does operator performance in condition L0 compare to that in condition L1? Several measures of task performance were used, on a per-task basis. Many of these are commonly used in traditional usability evaluations as well. E.g.,

- Mean time on task;
- Number of errors;
- Whether or not the user is in the process of transitioning from the use of one device to another;

- Or, more generally, the current phase of multi-device interaction.

Hypothesis

I hypothesize that operator performance measured via each of these metrics will be higher when there is a higher level of system support for task migration.

Approach

I attempted to detect differences in task performance metrics across L0 and L1 conditions for each task. The task-specific metrics that were used are described in detail in section §3.7.1, §3.7.2 & §3.7.3, for each task respectively.

1.3.3 RQ 3: Operator Performance and Subjective and Physiological Measures of Workload

Research Question

Are subjective assessments of mental workload an accurate indicator of operator performance in this domain? Are both, subjective measures of workload (NASA TLX) and the physiological measure (pupil radius), sensitive to workload in PIM tasks? It is clear that workload does not stay constant during a task, but varies constantly. What are the types of changes that can be observed in workload during the execution of a task? How do the two measures of workload each correlate with task performance?

Mental workload has been shown to be negatively correlated with several of these metrics in other domains [O'Donnell and Eggemeier, 1986, Ballas et al., 1992a, Bertram et al., 1992]. Does the same (or a similar) relationship hold between mental workload and task performance in the PIM domain?

Hypothesis

I hypothesize that changes in a few or all task performance metrics will be correlated with changes in mental workload. Thus, mental workload measured by NASA TLX can be used to predict operator performance in personal information management tasks. Subjective measures and physiological measures of workload will correlate with task performance metrics and with each other during the execution of a specific task.

Approach

To test this hypothesis, I obtained subjective ratings of workload and physiological measures (pupil radius) and attempted to correlate workload assessments with measures of operator performance.

1.4 Goals and Key Contributions

Via this work, I attempt to shed light on the differences in workload in PIM tasks, and their implications for research and practice.

1.4.1 Contributions to Research

This experiment was a first look at using mental workload measures in evaluating personal information management tasks. This research contributes to the field by examining changes in workload as users perform PIM tasks. It describes the role that subjective workload assessment scales such as NASA TLX and physiological measures can play as predictors of operator performance in these environments.

Very few differences were recorded in subjective assessments of mental workload between the two levels of support for each task, but significant differences were noted between different tasks. This suggests that while NASA TLX can discriminate between different tasks, it is not sensitive to changes within the execution of each task in this domain. The physiological metric, on the other hand, showed differences before and after the migration step for the Files task, as well as in all steps of the Contacts task. Since this metric highlights intra-task changes in workload that are not detected by subjective metrics, it appears to be a better choice for future workload studies in PIM tasks.

The lack of any meaningful correlation between performance-based metrics and workload metrics suggests that neither alone is sufficient to assess and describe highly contextualized tasks in the domain of personal information management. As has been noted elsewhere [Dillon, 2002a], traditional usability metrics focus on efficiency, effectiveness and satisfaction [International Standards Organization, 2008]. That constitutes the first paradigm of HCI [Harrison et al., 2007], stemming from its origins in the study of human performance and work practices. Users do not show particular concern for whether a common task takes a few seconds more or less to complete, but they do care about how the experience makes them feel: frustrated versus happy, weary and tired versus a joy to use. This research helps capture these subjective experiences, examines their relationship to traditional usability metrics, and identifies breakdowns in user activities caused by specific design factors [Bødker, 1989].

1.4.2 Contributions to Practice

Better and deeper knowledge of mental workload in information ecosystems can provide valuable formative feedback to designers, and assist them in creating systems that take these factors into account. Mental workload may be low throughout task performance on a single device, but exceptionally high at the point of transitioning from one device to another: this is hard to capture via traditional usability metrics. By simply measuring error rates, we may be overlooking the bigger picture: users may be trying hard to reduce errors, but in the process, incurring high mental workload.

The usability of a system incorporating multiple devices needs to be measured beyond traditional metrics such as task performance. Thus, this research can be used by designers to incorporate elements into their designs that actively aim to reduce mental workload for the operator. For example, systems such as Syncables [Tungare et al., 2007] were designed to reduce the workload on users by automatic support for migrating task-related data between two or among several devices in a user's computing environment.

1.5 A Guide to this Dissertation

This dissertation is organized as follows:

- This chapter presents an introduction to the problem domain, research questions, hypotheses and contributions of this dissertation to HCI research and practice.
- In Chapter 2, I situate my research within related prior work in Personal Information Management, Multi-Device User Interfaces and Mental Workload measurement.
- Chapter 3 describes in detail the two studies I conducted, including methodology, participant details, metrics used, and analyses performed. The first is a survey conducted to understand users' practices in PIM across multiple devices, and the second is a controlled laboratory study that explores this interaction in more detail.
- Chapter 4 presents the results obtained from the survey, and a re-examination of the research questions and hypotheses.
- Chapter 5 is a deeper discussion of some of the findings and implications of this research for the broader community of research and practice.
- Chapter 6 concludes with a summary of the work presented in this dissertation, and interesting questions that still remain unanswered and require future work.
- Appendices provide details of the experimental material used, IRB approval forms, analysis scripts, and other relevant material.

Chapter 2

Related Work

> "Where is the wisdom we have lost in knowledge?
> Where is the knowledge we have lost in information?"
>
> — *T. S. Eliot.*
> *In [Eliot, 1934]*

Managing information and being able to access it whenever and wherever necessary has been a concern for humankind since long before computers arrived on the scene. As new types and higher amounts of information are created and disseminated day after day, human attentional resources have stayed constant [Levy, 2005]. This has given rise to the problem of information overload [Schick et al., 1990]. The issue of information fragmentation across multiple devices threatens the effectiveness of users as well as of our tools and systems. An understanding of mental workload in PIM tasks is not only expected to lead to a better understanding of why a particular tool causes high frustration or mental demand in users, but also can be used to isolate critical sub-tasks and to assess the effectiveness of different tools.

In this chapter, I review prior work in the area of information management related to the topic of my dissertation and define some of the key terms to be used in later chapters. I will start from a general overview of information management, from before the age of computers. From there, I will proceed to a survey of the research in personal information management with computers and, later, portable devices. I continue with a closer look at the developments in designing interfaces for multiple devices. Finally, I discuss various ways of evaluating operator performance, including workload measures, and its use in the domain of personal information management across multiple devices.

2.1 Introduction

The idea and potential of digital information management can be ascribed to the vision of Vannevar Bush, from as early as 1945. In his seminal essay, "As we may think" [Bush, 1945], he laid the foundation for many influential ideas, many of which are only being realized recently. He described his vision of a Memex as a device *"in which an individual stores all his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility."* Many of Bush's ideas and propositions were realized in the decades that followed, though not in the exact form and technology he envisioned at the time. He is considered by many to be the visionary pioneer that inspired the field of digital information science [Fox et al., 1993] as well as the earliest proponent of a system that provides a personal memory record [Abowd and Mynatt, 2000]. At the time Bush first proposed his ideas, he referred to one's information as a *record* which would be *"continuously extended, must be stored, and above all, must be consulted"*. The term personal information management was not in common use until much later.

2.2 Personal Information Management

The earliest use of the term 'personal information management' can be traced back to an article by Mark Lansdale [Lansdale, 1988]. He refers to artifacts as *"personal information not necessarily in the sense that it is private, but that we have it for our own use. We own it, and would feel deprived if it were taken away."* Several definitions for Personal Information (PI) have been proposed. Bellotti et al. [Bellotti et al., 2002] define personal information management as *"the ordering of information through categorization, placement, or embellishment in a manner that makes it easier to retrieve when it is needed."*. Barreau [Barreau, 1995] identified five characteristics of personal information management systems: acquisitions, organization/storage, maintenance, retrieval and output. PIM is distinguished from general information management which involves non-subjective information that is managed collectively for/by more than one individual [Bergman et al., 2003].

A recent attempt at defining personal information by Jones [Jones, 2008] categorizes PI into six types. Personal Information is that which is:

1. **Controlled by or owned by us;**
   E.g. files, papers.

2. **About us;**
   E.g. medical information, tax records.

3. **Directed towards us;**
   E.g. email messages received, telephone calls.

4. **Sent (posted, provided) by us;**
   E.g. email messages sent, photos.

5. **(Already) experienced by us;**
   E.g. books read, web sites visited.

6. **Relevant (useful) to us;**
   E.g. advertisements, unread but related information.

2.2.1 Personal Information Management before Computers

As noted earlier, the seeds of personal information management were sown in 1945 [Bush, 1945]. Since managing one's information is intrinsically personal, a wide variety of practices can be observed among individuals. These practices are dictated by internal as well as external factors such as preference, context, training, etc. Many studies have investigated the nature of such individual personal information practices. Malone [Malone, 1983] conducted an exploratory observation of the information organization practices of users (by examining their offices and desks) before there was a computer on every desk. Lansdale [Lansdale, 1988] applied principles from psychology to explain some of the observed practices. He specifically discussed and explained some of Malone's findings in the light of psychological theory, and why his participants might have acted the way they did, based on their job requirements and several other factors. He examined how users categorized information, with the intent of simplifying the retrieval process later. Kwasnik [Kwasnik, 1989] studied the organization of documents in offices and factors that were important to users in classifying documents. She found that the use to which an item will eventually be put was one of the strongest factors in determining its classification.

In other domains that are also considered personal information, Kelley and Chapanis [Kelley and Chapanis, 1982] studied the use of paper calendars by professional persons. They discovered a wide diversity in the number of individual calendars maintained by their study participants, as well as a variety of archiving, accessing, and consulting patterns. They provided several guidelines from their study towards the expected computerization of appointment calendars. Payne [Payne, 1993] revisited calendars a decade after Kelley's and Chapanis's study; he found a similar diversity in the approaches to calendar-keeping. A few participants in his study used computers to maintain their calendars. The presence of digital copies of information has not caused users to discard their paper archives [Whittaker and Hirschberg, 2001, Tungare and Pérez-Quiñones, 2008a]; in fact, users still maintain highly-valued paper archives.

Since the advent of computers, several researchers have studied personal information management from various angles. In the physical domain, there used to be hardly any issues related to the same information existing in multiple places, as are common with digital information. Some of the more important issues today relate to accessing information from various sources, using different modalities and a variety of devices — problems which were not on the radar just a few decades ago.

As a prerequisite to my work, and to situate my work within the broader research agenda in

PIM, I present a detailed survey of the PIM literature.

2.2.2 Information Overload

Information overload is defined as occurring when the information processing demands on an individual's time to perform interactions and internal calculations exceed the supply or capacity of time available for such processing [Schick et al., 1990]. Information overload is often also referred to as information fatigue syndrome [Edmunds and Morris, 2000], information explosion [da Silva, 2005], information pollution [Nielsen, 2003], info glut [Denning, 2006], data smog [Shenk, 1998] and other terms that allude to negative environmental conditions for the knowledge worker. An excellent overview of the information overload problem is available in [Schick et al., 1990]. The problem has also been dealt with in detail by Edmunds and Morris [Edmunds and Morris, 2000]; they state that the inherent paradox is that the availability of vast amounts of information at our disposal has made it harder to locate the bits that we actually are interested in. Butcher [Butcher, 1995] identifies three dimensions of information overload: personal information overload, organizational information overload and customer information overload. Farhoomand and Drury [Farhoomand and Drury, 2002] conducted a study in which participants reported several meanings of the term 'information overload' as it applied to them: an excessive volume of information (79%), difficulty or impossibility of managing it (62%), irrelevance or unimportance of most of it (53%), lack of time to understand it (32%), and multiple sources of it (16%). Nelson [Nelson, 1994] explains the scale of the problem: *"more new information has been produced within the last three decades, than in the last five millennia. Over 9,000 periodicals are published in the United States each year, and almost 1,000 books are published daily around the world"*. The topic has also received significant coverage in the popular press as well. [1] [2]

As the problem of information overload has worsened over the years, human attentional resources have stayed constant [Levy, 2005]. In fact, they probably have decreased because of the increasingly busy lifestyles of today. The greater the volume of information, the more we spend our resources on determining if a particular piece of information is useful. Subsequently, we spend fewer resources on actually assimilating and using that information for productive work.

In the study I conducted, I simulated information overload conditions by presenting the participant/user with several independent tasks that required them to maintain some amount of state in their mind. Specifically, in the Files task (described in detail later in section §3.7.1), participants played the role of a consultant who worked with several clients. Each of these clients required the consultant to perform tasks for them, presented one at a time via an instruction display. In the Calendar task (section §3.7.2), I simulated a typical busy week for a family by including events that occurred both, at the office and at home, on weekdays and weekends, and during office hours and

[1] http://www.informationweek.com/551/51mtinf.htm
[2] http://www.infoworld.com/articles/ca/xml/00/01/10/000110caoverload.html

evenings. A few events were scheduled to overlap (intentionally). In the Contacts task (section §3.7.3), participants were provided the contact information of several new people who they 'met' at a conference, according to the script provided to them.

While information overload already creates a strain on the user, the situation gets worse when we take into account information fragmentation.

2.2.3 Information Fragmentation

Information fragmentation is the condition of having a user's data in different formats, distributed across multiple locations, manipulated by different applications, and residing in a generally disconnected manner [Bergman et al., 2006]. It is also referred to as *'compartmentalization of information'* [Bellotti and Smith, 2000].

Bergman et al. [Bergman et al., 2006] describe the case of information fragmentation for a chemistry student, Jane, who has her chemistry project-related data in three different formats under three different hierarchies: documents, emails, and bookmarks. There is a lack of any structural connection among the various format-related stores of information used to complete a single task, i.e. the file system, the browser bookmarks collection and the email inbox. A direct consequence of such information fragmentation is that when Jane needs to work on her chemistry project, she needs to use three different applications to deal with three different sources of information, each existing in a different format, with inherently different types, and situated in different contexts.

Bellotti et al. [Bellotti and Smith, 2000] describe a prototype PIM system that allows locating information irrespective of its format, according to user queries specifying required content or document properties. A few of the early users of this system used it to manage their email, documents, and notes in a common email-browser-like interface. The Stuff I've Seen system developed by Dumais et al. [Dumais et al., 2003] offers a similar cross-type interface to enable re-finding a user's personal information: email, web pages, documents, appointments, etc. The Haystack project [Adar et al., 1999, Huynh et al., 2002, Karger and Quan, 2004] had as one of its goals a type-agnostic approach to filing personal information. Commercial products such as Google Desktop [Google, Inc., 2004][3], Apple Spotlight [Apple, 2004] and Microsoft Windows Desktop Search [Microsoft, 2006] enable similar cross-type searches of users' data.

Bergman et al.'s definition of information fragmentation [Bergman et al., 2006] only included the fragmentation of information across different collections, e.g. files, email messages, and bookmarks all seemed to be managed within similar, yet duplicate, hierarchies [Boardman et al., 2003]. However, the issue of information fragmentation across multiple devices [Karger and Jones, 2006] looms larger as mainstream users increasingly have started to use portable devices such as cell phones, portable digital assistants (PDAs) and laptop computers for PIM [Tungare and Pérez-

[3]The author was an intern with the Google Desktop team in 2005.

Quiñones, 2008c].

The controlled experimental setup in my study incorporated information fragmentation across devices. The three tasks — file management, calendar management and contacts management — involved information spread over a desktop, a laptop, a phone, and paper.

2.2.4 Personal Information Collections

Bellotti and Smith [Bellotti and Smith, 2000] noted that information is managed in various collections independently, e.g. email, or documents, or bookmarks. [Boardman and Sasse, 2004] defines a collection as *"a self-contained set of items"* and that *"typically the members of a collection share a particular technological format and are accessed through a particular application"*. Information collections can vary greatly with respect to the number, form and content coherence of their items [Jones and Teevan, 2007]; it may or may not be strongly associated with a specific application. Many of the studies that have been conducted in the area of Personal Information Management (details in [Teevan et al., 2007]) are limited to how we manage information on a particular device (e.g. desktop), or how we manage a particular information collection (e.g. bookmarks or emails).

Among the studies focused on a single information collection in isolation are a few notable examples. In the study I conducted, I assigned tasks involving the first three of these collections.

- **Files.**

  Barreau and Nardi [Barreau, 1995, Barreau and Nardi, 1995] studied the contextual aspects of a person's work environment that guide the acquisition, classification, maintenance, and retrieval of documents. Their studies highlight that document attributes are not the only markers that guide PIM activities; that context plays an important role in users' decisions to keep and maintain their personal information collections.

- **Calendars.**

  Early research on calendar use predates electronic calendars. [Kelley and Chapanis, 1982] reported that the use of multiple calendars was prevalent, and a wide variation was seen in the time spans viewed, archiving practices, editing and portable access. Kincaid and Pierre [Kincaid et al., 1985] examined the use of paper and electronic calendars in two groups, and concluded that electronic calendars failed to provide several key features such as flexibility, power, and convenience, that paper calendars did at the time. Payne [Payne, 1993] theorized that the central task supported by calendars was prospective remembering (the use of memory for remembering to do things in the future, as different from retrospective memory functions such as recalling past events). A more detailed overview of calendar research is available in [Tungare and Pérez-Quiñones, 2008a].

- **Contacts.**

  Whittaker et al. [Whittaker et al., 2002a, Whittaker et al., 2002b] reported on issues related to contact management, stressing that this was a problem separate from email or communication management, although related. They describe the prevalent use of tools such as physical (paper) address books, digital address books, corporate directories, in-tool address lists, business cards, and sticky notes. Nardi [Nardi and O'Day, 2000] describe the design of a Contact Map that leverages contact importance and other social cues to visualize a user's social network.

- **Email.**

  Whittaker and Sidner [Whittaker and Sidner, 1996] noted that the increasing use of email for task management and personal archiving goes beyond its original purpose of communication, and that this causes email overload. Gwizdka [Gwizdka, 2000, Gwizdka, 2002, Gwizdka, 2004], Bellotti [Bellotti et al., 2003], Mackay [Mackay, 1988], Ducheneaut [Ducheneaut and Bellotti, 2001], and several others corroborate Whittaker's findings that email is being used to perform functions that email systems were not explicitly designed to handle. Each of them studied one of the many overloaded functions for which inboxes are being used. The increasingly-common use of email as a task management tool [Ducheneaut and Bellotti, 2001] has given rise to various strategies in users to manage this overload. Gwizdka [Gwizdka, 2004] examined the different management styles used for email and identified two groups of email users: the cleaners and the keepers. Stuff I've Seen [Dumais et al., 2003], Bifrost Inbox Organizer [Bälter and Sidner, 2002], Taskmaster [Bellotti et al., 2003] are some of the tools developed to assist email management.

- **Instant Messaging.**

  Instant messaging as a communication medium has not been widely studied yet, most likely because of its relatively recent popularity. Nardi [Nardi et al., 2000] conducted ethnographic studies of instant message (IM) use, and highlighted the uses of IM, i.e. negotiating availability, lighter-weight communication than using email, and presence awareness for distributed collaborative teams. It is also used as a preamble to other forms of communication such as using the telephone or email subsequent to an IM session.

- **Bookmarks.**

  Abrams [Abrams et al., 1998] studied the practice of users creating bookmarks to carve their own personal information space out as a subset of the entire World Wide Web. They probed the reasons behind creating bookmarks, how they were organized and maintained, and later retrieved. Jones et al. [Jones et al., 2002] studied the larger problem of how users organize web information for re-use (which involved bookmarking as well as various other techniques). Kelly and Teevan [Kelly and Teevan, 2003] performed longitudinal studies to understand users' web browsing behavior and relevance feedback.

2.2.5 Studies Spanning Multiple Information Collections

Bellotti and Smith [Bellotti and Smith, 2000] note the fragmentation of information into collections ("compartmentalization") due to poor integration of PIM tools. Amidst these narrow studies of specific collections, a notable exception is Boardman's cross-tool study of collections, [Boardman et al., 2003], which revealed similarities in the ways we manage disparate information collections. Attempts have been made [Chau et al., 2008] to leverage the relationships among data items in assorted information collections to build a graph of personal information items. This graph can then be used for multi-step searches, e.g. 'find the documents sent via email by the person I met at the meeting last Tuesday'. Teevan et al. [Teevan et al., 2004] studied users' strategies in locating their information, noting that users often navigated to their information in small steps (orienteering) instead of teleporting to it via tools such as search engines. Their study was more focused on information retrieval than organization and management, and was not restricted to personal information.

In my studies, I assessed the workload involved in performing tasks in three collections. Information and interaction in the three collection-based tasks do not overlap (the three tasks are performed in succession, not simultaneously), but the use of similar metrics for all three allows comparing them against each other.

2.2.6 Context in Personal Information Management

Context is an important factor in personal information management by virtue of it being related to a single individual or a group of individuals. Lansdale [Lansdale, 1988] reported that filing strategies and user-designed categories were highly contextual and exhibited large differences between users. This is one of the unique characteristics of personal information that differentiates it from general information management [Bergman et al., 2003]. [Gwizdka, 2006] proposed that that the contextual meta-data associated with personal information can be leveraged to assist information finding, keeping and organizing tasks. Referring to the dual problems of information fragmentation across collections [Boardman et al., 2003] and across devices [Tungare et al., 2006], Kirsh [Kirsh, 2006] identifies them as part of a larger issue, that of the distributed nature of context in PIM environments. He stresses that the notion of *place* in PIM is not a physical aspect, but organizational. Recent task-based approaches to PIM studied the use of task context to infer user action with the goal of learning users' PIM habits [Rath et al., 2008, Chirita et al., 2006]. Systems such as TaskVista [Bellotti and Thornton, 2006] and Project Planner [Jones et al., 2008] utilize activity- and project-related context to help users manage tasks, email, files, and other personal information efficiently.

I simulated the context-specific nature of PIM activities via specific instructions to participants that encouraged or forbade them from using certain devices in certain contexts. E.g. in the Files task (§3.7.1), participants were instructed that they were located either at their own office, or at a client's workplace. At each location, they could only use the device located at that (experimental) location, i.e., their desktop computer at their own office, and a laptop computer at the client's workplace. To reinforce the difference in context in both situations, each experimental location was associated with a physical location. Participants were requested to move physically to a different location when the instructions called for a change in location.

2.2.7 Re-finding Previously Encountered Information

Capra et al. [Capra et al., 2001] describe a system that allows a computer user to save contextual information associated with web browsing activities, and later to use them for accessing previously encountered information via a phone (voice). He presents [Capra, 2006] a detailed report of users' re-finding behavior. Dumais et al. [Dumais et al., 2003] designed a system, Stuff I've Seen, that captures a cross-collection log of information access by a specific user, and allows the user to search within this corpus for information already encountered in the past. Bergman et al. [Bergman et al., 2008] describe a user-subjective approach to PIM, and discuss the development of tools that de-emphasize the relevance of information accessed infrequently.

In my study, re-finding plays a relatively minor role. There are no re-finding-related instructions in the Files task. In the Calendar (§3.7.1) and Contacts (§3.7.3) tasks, later instructions require participants to lookup information they had entered/encountered in previous instructions.

2.2.8 Personal Information Management using Multiple Devices

PIM researchers recently have begun to examine the implications of the introduction of multiple devices into users' lives. The use of multiple devices raises novel issues, as has been pointed out earlier by us [Tungare and Pérez-Quiñones, 2008c] and others [Komninos et al., 2008]. A wide variety of devices is used, depending upon the context [Singh, 2006]. The mode of each device (silent/loud), form factor and other aspects all depend upon the context. This theme also has been the focus of a recent workshop dedicated to the investigation of personal information management off the desktop [Teevan and Jones, 2008]. Designs have been proposed for PIM devices of widely-varying form factors, from hand-held devices such as cell phones and PDAs [Robbins, 2008, Woerndl and Woehrl, 2008] to large table-top interfaces [Collins and Kay, 2008]. The need for light-weight information capture in mobile scenarios has been studied and well-documented [Bernstein et al., 2008].

In prior work I performed with Pyla et al. [Pyla et al., 2009], the issues that arise in multi-device interfaces, especially when several devices are used together to perform a single task, were described. We identified the lack of support for *task migration* in such interfaces, proposed the notion of *task continuity* across devices, described a system that implements a *continuous user interface (CUI)* and results from a preliminary investigation. The flow of information among a user's multiple devices has been likened to a biological ecosystem [Pérez-Quiñones et al., 2008]. Several concepts in Personal Information Ecosystems are analogues of related concepts from biological ecosystems, and the metaphor helps construct a meaningful information flow among devices. While task migration is handled at the interface level, seamless data migration requires system support. The Syncables framework [Tungare et al., 2006, Tungare et al., 2007] was developed in response to the need for being able to access data from any of a user's devices without extraneous task steps. Syncables assigns a unique human-readable address (a URI) to each data object and incorporates components that migrate Syncable data objects automatically and seamlessly among devices. Software that uses this framework needs only to request data items by URI from a well-defined connection endpoint (port) available on each device.

While I do not use the Syncables system in this study, it will be interesting to explore as part of future work, the impact of such a framework on mental workload for all three tasks that I studied.

2.2.9 Challenges in Studying Personal Information Management Practices

The unique idiosyncratic nature of personal information raises several challenges when researchers attempt either to study and categorize user behavior, or to evaluate new tools and systems with a diverse set of users, or to compare multiple tools against one another. Kelly [Kelly, 2006] and with Teevan [Kelly and Teevan, 2007] discuss these challenges in detail: PIM tasks are often performed at unpredictable times in response to a need for information; PIM tasks encompass a wide range of activities, not all of which can be effectively captured and studied; and laboratory studies tend not be reflective of users' personal information. The unique situational aspect of the working environment makes it difficult to study PIM as compared to general information storage and retrieval (ISAR) systems [Barreau, 1995].

Teevan et al. [Teevan et al., 2007] provide a categorization of types of studies performed in PIM; they include:

- **Interviews.**
  A majority of the studies reported in personal information management have relied upon in-depth personal interviews with participants to understand and gain insight into their information practices, followed by an analysis and categorization of the practices observed. In addition, a few researchers have provided cognitive or psychological analyses to explain their findings. The earliest interview-based studies in PIM were reported by Kelley and Chapanis [Kelley and Chapanis, 1982], Malone [Malone, 1983], Payne [Payne, 1993], and several others. For a deeper discussion, please see [Teevan et al., 2007].

- **Observational studies.**
  Jones et al. [Jones et al., 2001] conducted observational studies of participants at their workplace settings to understand their use of web bookmarks and other techniques for re-finding

web-based information. Observation was followed up with an interview with each participant; other instruments such as questionnaires and specific tasks were also included.

- **Surveys and questionnaires.**
  Because of the limitations of interviews in reaching a wider audience of participants, many researchers prefer to conduct survey-based studies to understand the practices of a larger set of users. Whittaker et al. [Whittaker and Hirschberg, 2001] surveyed 50 users to understand their paper archive management practices. Gwizdka et al. [Gwizdka, 2004] surveyed users about their email management practices. In Study 1, I conducted a survey of 220 participants about their information management practices across multiple devices [Tungare and Pérez-Quiñones, 2008b].

- **Log analyses.**
  Because of the methodological difficulties inherent in studying an activity such as PIM that happens on users' machines all the time, it is often difficult to gain a complete understanding of practices from one-time snapshots such as interviews or surveys. Other techniques such as diary studies or longitudinal studies are more appropriate for studying behaviors that are expected to change or evolve over longer periods of time. Log analyses of instrumented software are a good way to capture long-term information management patterns, e.g. Tauscher and Greenberg [Tauscher and Greenberg, 1997] studied the logs generated by 23 users to understand revisitation patterns for web sites.

- **Laboratory studies.**
  Laboratory studies yield specific types of data that cannot be obtained by any other means. When performance comparisons need to be made, or specific prototypes need to be evaluated, laboratory evaluations are the best choice. A limitation of laboratory studies applied to personal information management is that the underlying information structure is not familiar to each participant. There is a tradeoff involved between experimental control and the "personal-ness" of the information used in the study. E.g. Kaasten et al. [Kaasten et al., 2002] conducted a controlled study to understand users' revisitation patterns for web sites; Capra et al. [Capra and Pérez-Quiñones, 2004] conducted a laboratory study involving a collaborative dialog with another participant to examine the use of shared context in information re-finding.

Each method has its advantages and limitations; the issues of user interaction across multiple devices that I wanted to study are not amenable to study via either interviews, surveys or log analyses. Interviews capture users' understanding of the situation, but the data obtained via self-reports fails to capture the changes in workload during the performance of the task. Surveys suffer from the same limitations, being self-reports. Log analyses, while useful in providing post-facto evidence of certain phenomena, are not particularly useful when the primary interest is in interaction rather than information. While observational studies and laboratory experiments both showed promise in studying multi-device personal information management, the lack of a controlled environment in observational studies was considered a serious limitation. Because of all these factors, I conducted a controlled laboratory experiment, described in chapter 3.

2.3 Multi-Device User Interfaces

One of the major causes of information fragmentation is that we no longer are restricted to a single device, or a single source of information; most of our information is scattered across multiple devices, such as desktop computers at the office, laptops at home, portable digital assistants (PDAs) on the road, and of course, cellphones. These physical manifestations echo the "computing by the inch, the foot and the yard" research at Xerox PARC initiated by Mark Weiser [Weiser, 1991]. Computation is no longer confined to the desktop, just as he outlined a decade and a half ago [Weiser, 1994].

2.3.1 Interaction in a Mobile Context

It has widely been recognized that the mobile context is fundamentally different from the stationary context, and design must therefore account for the differences [Perry et al., 2001, Oquist et al., 2004]. Indeed, mobile interaction occurs in an entirely different "place" [Harrison and Dourish, 1996] than desktop-chained interaction. Dourish refers to situated interaction as *"embodied interaction"*, and outlines several principles that designers must take into account for technology that, by its very nature, must co-exist in the environment that users use it in. Not only does mobile interaction happen in a different context, it also places different attentional demands on the user [Oulasvirta et al., 2005]. Attention in mobile contexts is fragmented and users devote as few as 4 seconds at a time to the task at hand. My experiment helps understand the nature of this fragmented attention at the point of transition from one device to another.

2.3.2 Interface Adaptation and Migration

Much of the work in multi-device interfaces has focused on adapting an interface designed for one device automatically to another device. There is a wide variety of approaches in performing automatic interface translation. Prior work has also explored a variety of approaches to building multi-platform user interfaces: these include model-based approaches [Mori et al., 2003, Einsenstein et al., 2001, Thevenin and Coutaz, 1999], Interface Description Language (IDL)-based approaches [Abrams et al., 1999], UsiXML[4], XForms[5], XUL[6] and XAML[7], etc., transformation-based approaches [Richter, 2005, Florins and Vanderdonckt, 2004] and task migration techniques [Johanson et al., 2001, Chu et al., 2004, Chhatpar and Pérez-Quiñones, 2003].

2.4 Holistic Usability in Multi-Device Environments

Computing does not occur as a stand-alone dedicated task, nor does it happen in a predefined, dedicated window of time. We live in a complex networked world of devices, and use several devices together, and much of our knowledge is in the world [Norman, 1988]. The origins of the art and science of usability and human factors can be traced back to factories and environments where users performed specific duties at specific times. The goal of the human factors specialists was to optimize operator performance and the fit between human and machine; that is the first paradigm of HCI [Harrison et al., 2007].

The definition of usability, according to the International Standards Organization [International Standards Organization, 2008] (*"the effectiveness, efficiency and satisfaction with which specified users can achieve specific goals in particular environments"*) includes guidance on the design of visual displays, input devices, accessibility, etc., but fails to include guidance on designing for situated use in a computing environment. It thus establishes a necessary, but not sufficient, condition for product quality.

Modern developments in the science of cognition have examined the relationship of the user in complex computing environments, and place greater emphases on the situational aspects of human-computer interactions. Distributed cognition theory [Hutchins, 1995] extends the reach of what is considered cognitive beyond the individual to encompass interactions between people and with resources and materials in the environment. As Hutchins stresses, *"the proper unit of [cognitive] analysis is, thus, not bounded by the skin or the skull. It includes the socio-material environment of the person, and the boundaries of the system may shift during the course of activity."* In multi-device computing environments, it is worthwhile to analyze the system as an integrated whole whose purpose is to assist the user in satisfying her information needs (also the definition of a Personal Information Ecosystem, in prior work I performed with Pérez-Quiñones, Pyla and Harrison [Pérez-Quiñones et al., 2008]). Taking into account such a holistic perspective, we may better be able to understand the processes that take place in the system and how they may be supported, enhanced and augmented with external cognitive aids. Hollan notes [Hollan et al., 2000] that such systems are defined by their role and function, not necessary location; thus, a user's mobile devices are an integral component of her personal information ecosystem even when they are not collocated.

[4] http://usixml.org/
[5] http://www.w3.org/TR/xforms/
[6] http://www.mozilla.org/projects/xul/
[7] http://msdn.microsoft.com/en-us/library/ms752059.aspx

Other recent theories such as Embodied Interaction [Dourish, 2001] also support the notion that technology and practice are closely intertwined; they co-exist and co-evolve. Recognizing and exploiting the richness of this interaction allows for better support of situated tasks. Dourish notes the artificial boundaries imposed by devices and interfaces on a user's tasks [Dourish, 2001] (pg. 198–199). Better designs should allow for user-initiated renegotiation of these arbitrary boundaries, paving the way for an information-rich environment that is not torn apart in the purported paradox between device convergence and information appliances. He stresses that interaction designers no longer design interfaces, they design experiences. (pg. 202). In particular, I note that today's interfaces for multi-device interaction do not do a good job of translating a user's action into meaning. An interface that does respect such intentionality would provide (semi-)automatic support for task migration when a user is detected as having changed focus from one device to another, making the latter worthy of his dominant attention. Harrison, Tatar and Sengers [Harrison et al., 2007] refer to these as non-task-oriented computing activities. The name reflects the property of these interfaces to stay invisible [Norman, 1999], thus not amenable to evaluation using traditional performance metrics.

2.4.1 Hot Cognition Aspects in the Evaluation of Personal Information Ecosystems

Norman [Norman, 2003] argues that emotion plays a central role in our interaction and appreciation of the computing devices we use. But classic usability metrics fail to account for subjective factors such as emotional appeal, frustration, and likability. All these point to the necessity of bringing hot cognition aspects into the evaluation process. Other approaches stress the inclusion of non-mainstream aspects in design: Jordan [Jordan, 2000] advocates designing for pleasurability of the user. Deriving from Maslow's hierarchy of needs [Maslow, 1943], Jordan identifies a hierarchy of needs for a computing system: functionality is the most basic, level 1. The next level up is usability, and beyond that — level 3 — is pleasure. In hierarchy, therefore, usability is necessary but not sufficient to guarantee an optimal user experience. Layard [Layard, 2006] reports on his investigations of 'happiness' as a psychological metric that merits scientific study.

Kelly et al. [Kelly and Teevan, 2007] identify a shortcoming in PIM studies as well; quality of life measures, such as those developed by [Endicott et al., 1993] have received received little attention in PIM evaluations, despite the goal of the field being to make people's lives easier. Bagozzi [Bagozzi, 1992] suggested that researchers have not been able to predict human behavior with simply cognition and affect because the construct of conation has been absent from this analysis. Conation is the motivation or strong desire of a person to take certain actions based on the one's current cognitive and affective state. It is interesting to study why users act the way they do in their interaction with devices. An expanded definition of usability clearly can broaden the search for these factors, and hopefully, explain the nature of distributed usability. There is opportunity for the introduction of pleasurability scales, such as the ones developed by Davis [Davis, 1989]: Perceived Usefulness and Perceived Ease of Use as predictors of user acceptance. The importance of Perceived Ease of Use in user acceptance is in agreement with Bandura's theory of self-efficacy [Bandura, 2000].

2.5 Mental Workload Assessment

Several PIM studies describe in detail how information may be captured, organized, archived and accessed for fast retrieval, high relevance, and efficient task processing. Certain aspects of memory have been examined in PIM [Elsweiler et al., 2006]; the related, yet distinct, issue of mental workload needs deeper study because it can help understand issues that are not captured by traditional usability metrics alone [Dillon, 2002a]. Especially, aspects such as frustration and measures of perceived performance and effort are expected to contribute to an understanding of users' frustration regarding current PIM tools [Tungare and Pérez-Quiñones, 2008b].

Mental workload is defined as *"that portion of operator information processing capacity or resources that is actually required to meet system demands"* [O'Donnell and Eggemeier, 1986, Eggemeier et al., 1991], or *"the difference between the cognitive demands of a particular job or task, and the operator's attention resources"* [Wickens, 1992]. It is task-specific and operator-specific (i.e., person-specific) [Rouse et al., 1993]; the same task may evoke different levels of workload in different individuals. *Task complexity* is related to the demands placed on an operator by a task, and is considered operator-independent, whereas *task difficulty* is an operator-dependent measure of perceived effort and resources required to attain task goals [de Waard, 1996]. Mental workload is considered an important, practically relevant, and measurable entity [Hart and Staveland, 1988].

2.5.1 Measures of Mental Workload

Eggemeier et al. [Eggemeier et al., 1991] classify workload assessment techniques into one of three categories:

- Performance-based assessment techniques,
- Subjective workload assessment techniques, and
- Physiological workload assessment techniques.

Muckler and Seven [Muckler and Seven, 1992] and de Waard [de Waard, 1996] suggest the use of the term *self-report measures* instead of *subjective measures* to reflect the fact that physiological measures are also sometimes subjective.

Each of these differ along several dimensions: [Eggemeier, 1988, Eggemeier et al., 1991] identify six important properties of workload assessment scales:

- Sensitivity,

- Diagnosticity,
- Intrusiveness,
- Reliability,
- Implementation requirements and
- Operator acceptance.

Among the many types of measures, subjective measures are becoming an increasingly important tool in system evaluations and have been used extensively to assess operator workload [Rubio et al., 2004] because of their practical advantages (ease of implementation, non-intrusiveness) and their capability to provide sensitive measures of operator load. Their significant advantages over performance-based assessment techniques make them the preferred candidate in task domains where instrumenting the equipment for the primary task is expensive/difficult.

2.5.2 Performance-based Assessment Techniques

Performance-based assessment techniques assess workload by measuring the operator's capability in specific task scenarios. In computer-based tasks, these include metrics such as the number of errors committed in task performance, time on task, and many other task-dependent measures.

By definition, performance-based metrics are highly task-specific, and researchers must determine the applicable metrics for every task context independently. In several domains, obtaining accurate measures of task performance requires special instrumentation of the equipment used. Often, such instrumentation may be expensive to attempt, or may change the fundamental nature of the task. Hence, task performance measures are used only in cases where no other measures may provide reasonably accurate measurements, or if such instrumentation is relatively cheap/easy.

I discuss the task performance measures used in the experiments I conducted in sections §3.7.1, §3.7.2 and §3.7.3.

2.5.3 Subjective Workload Assessment Techniques

Subjective mental workload assessment can be defined as the subject's direct estimate or comparative judgment of the mental or cognitive workload experienced at a given moment [Reid and Nygren, 1988]. It has been reported that although subjects may not be able to observe their own cognitive processes directly, they still may be able to report accurately about them [Nisbett and Wilson, 1977]. Several rating scales for workload assessment have been developed, presented here chronologically:

- Cooper-Harper Scale [Cooper and Harper, 1969];
- Modified Cooper-Harper (MCH) Scale [Wierwille and Casali, 1983];
- Subjective Workload Assessment Technique (SWAT) [Reid and Nygren, 1988];

- NASA Task Load Index (TLX) [Hart and Staveland, 1988];
- Workload Profile (WP) [Tsang and Velazquez, 1996].

Their various properties have been evaluated in a wide range of application domains [Bertram et al., 1992, Ballas et al., 1992b, Schryver, 1994]

The NASA Task Load Index (TLX)

The NASA Task Load Index (NASA TLX) [Hart and Staveland, 1988] is a multi-dimensional subjective workload assessment technique, developed as a measure of perceived workload. In the years since, it has been shown to be a highly reliable, sensitive measure of workload [Hendy et al., 1993, Rubio et al., 2004]. It has been applied in studies of airline cockpits [Ballas et al., 1992a], navigation [Schryver, 1994], in the medical field [Bertram et al., 1992], and in several other task scenarios. It includes six bipolar dimensions, as summarized in the table in Appendix §7.6. It combines information about specific sources of workload weighted by their relevance, thus reducing the influence of those that are experimentally irrelevant, and emphasizing the contributions of others that are experimentally relevant. This reduces between-subjects variability for the measure as compared to other subjective scales.

Rubio et al. [Rubio et al., 2004] compared three scales, NASA TLX, SWAT and WP on various dimensions; they found that the concurrent validity (as examined by the degree of agreement between the subjective workload and performance measures) was highest in TLX as compared to the other two. SWAT and WP showed lower correlations with performance. [Battiste and Bortolussi, 1988] determined that, between TLX and SWAT, TLX proved sensitive to a few mental workload differences not discriminated by SWAT [Rubio et al., 2004]. Hill et al. [Hill et al., 1989] also found that TLX had the highest sensitivity among the four scales (TLX, WP, SWAT and OW). Since my experiment is concerned with using a subjective workload assessment measure to predict task performance, the choice was made to use NASA TLX.

However, while the NASA TLX is effective as a per-task measure of workload, it is a post-facto measure that is responsive only to the overall workload during the course of the task. It is not a continuous measure and thus cannot be used to determine workload during sub-tasks with higher workload requirements. In my experiment, I needed a continuous measure of workload to be able to identify issues resulting from the transition between devices. Clearly, subjective measures are not sufficient to provide data at this granularity, hence I also used physiological workload assessment techniques.

2.5.4 Physiological Workload Assessment Techniques

Several physiological changes occur in response to variation in mental workload: these include changes in electro-encephalographic activity (EEG) [Schacter, 1977], event-related brain potentials (ERP) [Kok, 1997], magnetic field activity (MEG), positron emission tomography (PET), electro-oculographic activity (EOG) [Kramer, 1991] and pupillometric measures [Beatty, 1982, Backs and Walrath, 1992, Granholm et al., 1996]. Their measurement requires specialized equipment such as amplifiers, trackers, transducers, cameras, large storage media, etc., which make them substantially more expensive than other measures. Some physiological measures cannot be measured within an adequate time interval after the principal stimulus has been administered, e.g. changes in hormonal activity. While standardized scoring procedures have been developed for subjective and task-based measures, the interpretation of physiological data requires an extensive amount of technical expertise [Kramer, 1991].

Among their advantages is the possibility to provide a continuous measure (as opposed to subjective measures which provide a task-level estimate, not a continuous estimate). They do not introduce any extra steps in the tasks that are performed, and therefore are unobtrusive (although their measurement requires specialized probes and trackers to be worn by the subject).

Pupillometric Measures

Hess [Hess, 1975] speaks of merchants from several centuries ago who observed clients' eyes to infer interest in a product — this is the earliest reported relationship between pupil radius and attention. The topic has been dealt with with modern scientific rigor by several researchers since [Hess and Polt, 1964, Kahneman, 1973, Beatty, 1982, Klingner et al., 2008]. The observed changes in pupil radius in response to specific task-related changes in information processing are referred to as the Task-Evoked Pupillary Response (TEPR) [Beatty, 1982, Klingner et al., 2008]. Measurement of pupil radius is sensitive, real-time [Beatty, 1982, Kramer, 1991, Klingner et al., 2008], and easy to administer using eye tracking equipment.

TEPR has been used as a physiological measure of mental workload in several studies [Iqbal et al., 2005, Bailey and Iqbal, 2008]. Both, head-mounted eye trackers [Marshall, 2002] and desktop eye trackers [Klingner et al., 2008], can be used to measure task-evoked pupillary response. Within a single task, mental workload decreases at sub-task boundaries [Iqbal and Bailey, 2005]. Such continuous measures of mental workload can help locate sub-tasks of high task difficulty.

In addition, pupillary response is rapid, usually within 600ms of an eliciting stimulus [Kramer, 1991] and thus is effective as a continuous, near-real-time estimate of mental workload [Wilson and Eggemeier, 1991]. Kramer argues [Kramer, 1991] that although there have been reports of failures in obtaining a systematic relationship between pupil diameter and task difficulty [Wierwille et al., 1985], the specific studies contained several methodological deficiencies which could have been the source of the observed discrepancy. Negative results also have been reported by [Schultheis and Jameson, 2004], but these appear to be in the minority.

As mentioned earlier, the property of pupillometric measures to provide a reliable, continuous, quick and easily-instrumentable measure of workload made it a good choice in my study to assess workload levels at the point of transition between devices.

2.5.5 Using Multiple Assessment Techniques

Although directly measured, performance metrics are unique to a particular task, and there are several reasons why performance metrics cannot be used to predict performance for an unknown task [Wilson and Eggemeier, 2006]. Wilson and Eggemeier [Wilson and Eggemeier, 1991] recommend the use of multiple assessment techniques since no single technique is adequate in all situations. High correlation has been found between subjective measures and physiological measures in several studies [Roscoe, 1984, Speyer et al., 1988]. Since the ratings obtained via subjective workload assessments are not task-specific, it is possible to use them to compare the workload imposed by different tasks. Several task performance situations involving computing devices in the environment have been examined in detail, and workload measurements have been conducted using some of the techniques listed above [Ballas et al., 1992a, Schryver, 1994, Bertram et al., 1992].

2.6 Summary

In this chapter, I presented a detailed review of prior related work in personal information management, multi-device user interfaces and the measurement of mental workload. My work extends prior research in ways that seek to fill the gap among these areas. The contribution of this dissertation is that it examines personal information management across multiple devices, especially at the point of transition between them.

Chapter 3

Methodology & Analysis

3.1 Introduction

This chapter describes the experimental methodology used to study the research questions described in section §1.3. To do this, I conducted a controlled experiment with users performing PIM tasks on multiple devices. Instead of picking experimental tasks by making inferences or assumptions based on existing literature, I conducted a survey to elicit these from users themselves. Here I describe the details of the survey first (§3.2), followed by the experiment design (§3.6), a list of experimental tasks (§3.7), and measures used during the experiment (§3.4).

3.2 Study 1: Exploratory Survey Study

While trying to develop a set of tasks for the experiment, I realized that current literature on personal information management and multiple devices did not provide any background information on what such tasks could be. My personal experience and conversations with several colleagues (students, professors, friends) revealed that multiple devices are used together in an ad hoc manner, and this usage is very idiosyncratic. This movement of information from one device to another had not been captured in any study reported.

Representative tasks to be used for the experimental portion of this study needed to be frequent, critical, and real [Whittaker et al., 2000]. The best set of tasks thus would be those that are performed by several real users many times a day as a critical part of their work. Thus, in order to elicit these tasks, I decided to conduct a survey of knowledge workers, asking them several questions about their common tasks, devices, problems, solutions and outcomes.

The first study was strictly exploratory: the objective was to gain knowledge about the state-of-the-art in the field, to understand users' current set of devices, activities, problems, solutions, etc. The research questions (below) reflect this exploratory nature; there were no premeditated hypotheses to be confirmed or rejected.

3.2.1 Research Questions

The chief research questions of interest for this study were as follows:

1. What devices do users commonly use?
2. Which among them are used together with one another?
3. How do users adapt their workflows to their devices?
4. What problems do they encounter and how do they avoid them?
5. What strategies evolve and what role does mental workload play in these strategies?

3.2.2 Survey Design

In August 2007, based on the above general questions, I conducted a survey to understand the information management practices of users who use more than one information device. Since we wanted to study the usage patterns of users who used multiple devices to varying degrees, we concentrated on the population that was most likely to use many such devices in their everyday life: our audience largely consisted of knowledge workers, including professionals, students, professors, and administrative personnel. The survey was administered via the Internet to be able to reach beyond just the local population, and received responses from participants working at several companies in the San Francisco Bay Area (Google, Apple, IBM, Yahoo!), many universities (Virginia Tech, Georgia Tech, Michigan State University, Bath University UK) and companies based in Bangalore, India. ($N = 220$). While other experimental techniques such as ethnographic studies, personal interviews, and guided tours would have provided richer data about fewer participants, the intent behind this survey was to catalog a wide range of experiences.

The survey contained a mix of quantitative and qualitative questions, and a preliminary analysis was reported in [Tungare and Pérez-Quiñones, 2008b]. The questionnaire used for the survey (along with the IRB approval form) is available as an appendix (Sections §7.1, §7.2). It was administered using a web application hosted by Stellar Survey[1]. While Virginia Tech's in-house survey administration tool, survey.vt.edu[2] was considered, it did not support several features I needed; specifically, the ability to group questions in tables, and to allow for more forms of interaction such as configurable drop-down boxes.

Questions belonged in the following four categories:

[1] http://stellarsurvey.com
[2] http://survey.vt.edu/

Devices and Activities

- What is the distribution of users who use multiple devices?
- Is it only a small fraction of the population, or a larger majority?
- What are a few of the most common devices?
- What are some of the common PIM tasks that users choose to perform on certain devices?
- Are there certain tasks that are bound to a particular device, such that they may only be performed on that device?
- Are there certain tasks that may never be performed on certain devices?

The Use of Multiple Devices Together

- Which devices are commonly used in groups (i.e. together with other devices, used either simultaneously, or one after the other) to perform common tasks?
- What are the methods employed to share data among these devices?
- Do users keep their grouped devices completely synchronized at all times (i.e. do they maintain a copy of the same data on both devices at all times)?
- What are some of the problems and frustrations users have faced in using multiple devices together?
- Are people completely happy with the current offerings and their own workflows, or are they frustrated by certain aspects of how they are forced to manage their information by the current crop of tools and systems?

Buying New Devices

- What are some of the factors that influence users' buying decisions for new devices?
- How important it is to them that a new device integrate well into the set of existing devices?

Device and System Failures

- How often do users encounter failures in their information management systems?
- What are some of the common types of failures?
- How do users cope with failures?
- Are there any systems in place to guard against such failures, or are there reliable means of recovery from failures, after they occur?

3.3 Analysis of Study 1

A complete analysis of the data collected in Study 1 (survey) guided the experiment design for Study 2. The analysis was performed in two parts: a quantitative analysis of the questions regarding users' use of devices and their activities; and a content analysis of the free-form responses to identify some of the tasks and operations that were reported as frustrating or difficult to perform in their day-to-day usage. Three of the most-often mentioned tasks from the content analysis were used as representative experimental tasks for Study 2.

3.3.1 Content Analysis Procedures

Figure 3.1 shows an example comment from one of our survey participants. In figure 3.2, important elements of this comment are highlighted and tagged. Elements of interest were pooled from all participants' responses, and they enabled the design of representative experimental tasks, including the specific devices used, their location, context, information stored on them, and features (or lack of features.)

> "The last device I acquired was a cell phone from Verizon. I would have liked to synchronize data from my laptop or my PDA with it but there seems to be no reasonable way to do so. I found a program that claimed to be able to break in over bluetooth but it required a fair amount of guess work as to data rates etc and I was never able to actually get it to do anything. In the end I gave up. Fortunately I dont know that many people and I usually have my PDA with me so it isnt a big deal but frankly I dont know how Verizon continues to survive with the business set..."

Figure 3.1: Example comment from survey participant

3.3.2 Tag Types

From a preliminary examination of the data, five types of elements of interest (hereon referred to as 'tags') were evident. These were decided **a priori**, while the set of actual tags in each type were subject to **emergent coding**. Since the content to be analyzed was in response to specific questions, there was very little variation in terms of the units of information present in each response. The five tag types as outlined below were pre-decided before actual coding begun. Tags were of the following five types:

- `device`:
  Device(s) reported by the participants that were used (successfully or unsuccessfully) in performing the task. If a specific device or type of device was mentioned, then this tag was

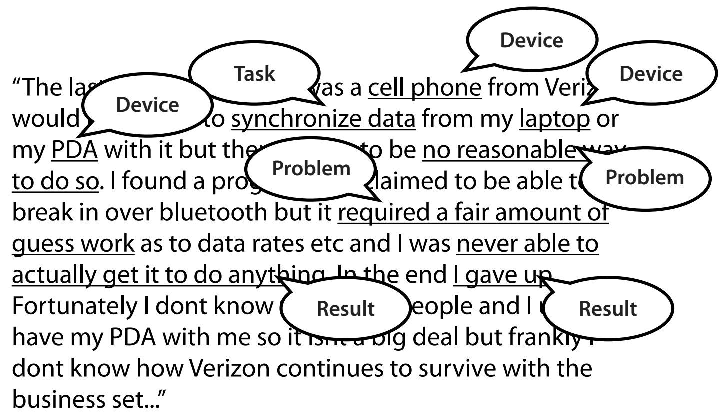


Figure 3.2: Tagging and analysis of example comment

applied. The value of the tag indicated which of several types of devices was mentioned. Multiple instances of this tag were allowed per unit of analysis.

- `task`:
  The specific task that the user was trying to perform. In many cases, users reported more than one task.

- `problem`:
  The problem encountered by the user while trying to perform the task. Several problems might be reported by each user per task.

- `solution`:
  Solutions that the users came up with, in order to perform the task. Some of these were workarounds developed in response to the fact that the existing software and solutions did not satisfy users' needs.

- `result`:
  The final result of the users' efforts — whether or not they were successful in performing their tasks.

Tagging was done using steps described in [Krippendorff, 2004]. Each response was treated as an independent sampling unit as well as a context unit. Due to the short length of each response (typically from one to ten sentences), there was no need to establish a shorter context unit. The recording unit was the actual datum provided by the user in either of the five categories of tags: either a device, a task, a problem, a solution, or a result.

3.3.3 Tags

Commonly-occurring themes were semantically judged, and a new tag was established for each new concept. Table 4.1 shows the Tag Types that were defined, and the tags in each tag type. The intention was to express each tag as a succinct phrase that captured the details of the relevant device, task, problem, solution or result. E.g. `task:browseTheWeb`, `problem:backupFailed`, `problem:conflictingEdits`, etc. (A complete list is presented in Table 4.1.)

The purpose of this preliminary analysis was to establish a set of representative tasks, devices, and contexts for Study 2. Because these results were not intended to be reported widely beyond their prescriptive use in Study 2, content analysis was performed by the experimenter alone. Two passes were conducted over the coded data.

During the first pass of coding, several ideas were found to have been expressed as slightly different wording in the tags (e.g. differing only in lower case or upper case, or the insertion of superfluous adjectives or adverbs). Such obvious duplicates were merged by replacing all occurrences with a single canonical form of the semantic significance of each tag. E.g. `task:browseTheWeb` and `task:browsingTheWeb` were collapsed into one, `problem:conflictingEdits` and `problem:editConflict` were collapsed as well.

Statistical analyses of the quantitative response and a content analysis of the responses to open-ended qualitative questions are presented in §3.3.

3.4 Study 2: Experimental Measurement of Mental Workload

From the results of the survey (reported in §3.3), users consistently reported difficulties in performing information management tasks with multiple devices, especially when transitioning between/among devices. From the responses received, I determined (from a content analysis of free-form responses) that users' adoption of various technological alternatives was guided by an innate sense of certain specific factors. I noted that several of these factors constitute mental workload, e.g. frustration level, temporal demand, and mental effort.

3.4.1 Abbreviations and Terminology

In several places in the rest of this dissertation, I refer to short codes, consisting of a letter followed by a number. Here is a detailed explanation for the guidance of the reader:

Participants (P1–P21)

Participants, wherever specifically referenced, are prefixed with '**P**'. E.g. P1, P2, etc.

Sessions (S1, S2)

The experiment was conducted in two sessions (details in section §3.6). I refer to the first session as S1, and the second session as S2.

Tasks (T1–T3)

There were three experimental tasks (§3.7), referenced as below:

- T1: Files (§3.7.1)
- T2: Calendar (§3.7.2)
- T3: Contacts (§3.7.3)

Level of System Support (L0, L1)

Participants performed each task at two levels of system support:

- L0: Lower level of system support for migrating information across multiple devices,
- L1: Higher level of system support for migrating information (than in L0).

Task Instructions (I1–I17)

For each task, users were provided a set of instructions one after the other. I prefix each of these instructions with 'I', e.g. I0, I1, etc.

3.5 Representative Tasks from Survey

In order to answer the research questions outlined in sections §1.3.1, §1.3.2 & §1.3.3, I designed an experiment consisting of three different experimental tasks, each at two treatment levels. From a content analysis of survey data from Study 1 (§3.3), the following emerged as the most common tasks where users encountered problems.

3.5.1 File Synchronization

One of the most commonly reported frustrating tasks that emerged was synchronizing data (this echoes findings by others [Dearman and Pierce, 2008]). Users' responses to this question elicited a long list of problems and issues that they often encountered. Because an overwhelming majority of respondents expressed difficulty and frustration at this common task, I assigned accessing files from multiple devices as one of the experimental tasks to our participants.

Participants indicated that they often used USB drives to bridge the gaps in their information management workflows when using multiple devices (n=6) as part of their regular work. In addition to these, a few participants indicated that they used USB drives as a workaround when their regular information management strategies did not work (n=5). It was interesting to study the use of these storage devices which were not designed explicitly for making conflict merging easier, but have been repurposed by a significant portion of the population to that end.

Since both these tasks involve file management in different ways, they were merged into a single experimental task: the **Files Task**.

3.5.2 Accessing and Managing Calendars

Many participants indicated that they had trouble accessing their calendars across multiple devices. In terms of numbers reported, this was ranked at #2, just below the data synchronization task. One of the main motivations for using more than one device was to be able to access their calendar information when away from their desks. The use of paper calendars is widespread, even despite the availability of online calendars. It is not clear which of these methods is easier; almost equal numbers of participants reported preferring one over the other for several reasons (details are available in [Tungare and Pérez-Quiñones, 2008a]).

Thus, the **Calendar Task** was chosen as the second experimental task.

3.5.3 Using a Phone to Manage Contacts

Phones are used by many users for common tasks such as making phone calls, sending messages, and increasingly, accessing information. A secondary task that needed to be performed was synchronization of information between their computers and their phones. Many of these tasks caused problems. Our participants commonly identified these as related to deficiencies in the phone interface (n=5), or a lack of features in the specific software they used, both on the computer as well as on the phone (n=3). They also cited that their phone contained many features which were of no use to them, but only served to complicate the interface (n=3) or that the system entirely crashed (either the phone software or the synchronization software, when the task involved such synchronization.)

The third task in my experiment was the **Contacts Task**, where participants were required to use a phone and a laptop for contact management, with and without system support for synchronization.

3.6 Experiment Design

In this experiment, I was interested in the impact of two factors — task, and level of support — on workload in participants. Since individual differences in work practices, task performance, and assessments of workload would display high variability across participants, a within-subjects design was used. Each participant constituted one experimental unit; each participant was assigned to each treatment level, and performed all three tasks at both levels.

Each participant was assigned to each cell, making this a complete block design (at 3×2 treatment levels). Each experimental task identified above was assigned to users to be performed in one of two sessions separated by at least two weeks, in order to minimize the learning effects associated with the first session.

The order of tasks assigned within a session was completely counterbalanced. Since there were 3 tasks, equal numbers of participants were assigned those tasks in the orders ABC, ACB, BAC, BCA, CAB, CBA (6 each, for a total of 18 participants $\times$ 2 sessions; data from 3 participants collected during session S1 was dropped (details in §4.2.1). The assignment of levels of support (L0 and L1) to participants during each session was subject to incomplete counterbalancing. Since L0 (lower level of system support for information migration) was deemed of higher task difficulty than L1 for all three tasks, no participant was assigned all three L0 (or L1) tasks during a single session. Each participant was assigned two L0 (or L1) tasks, and at least one L1 (or L0) task.

Figure 3.3 shows a graphical overview of the entire experimental setup. A detailed description of experimental tasks follows in section §3.7.

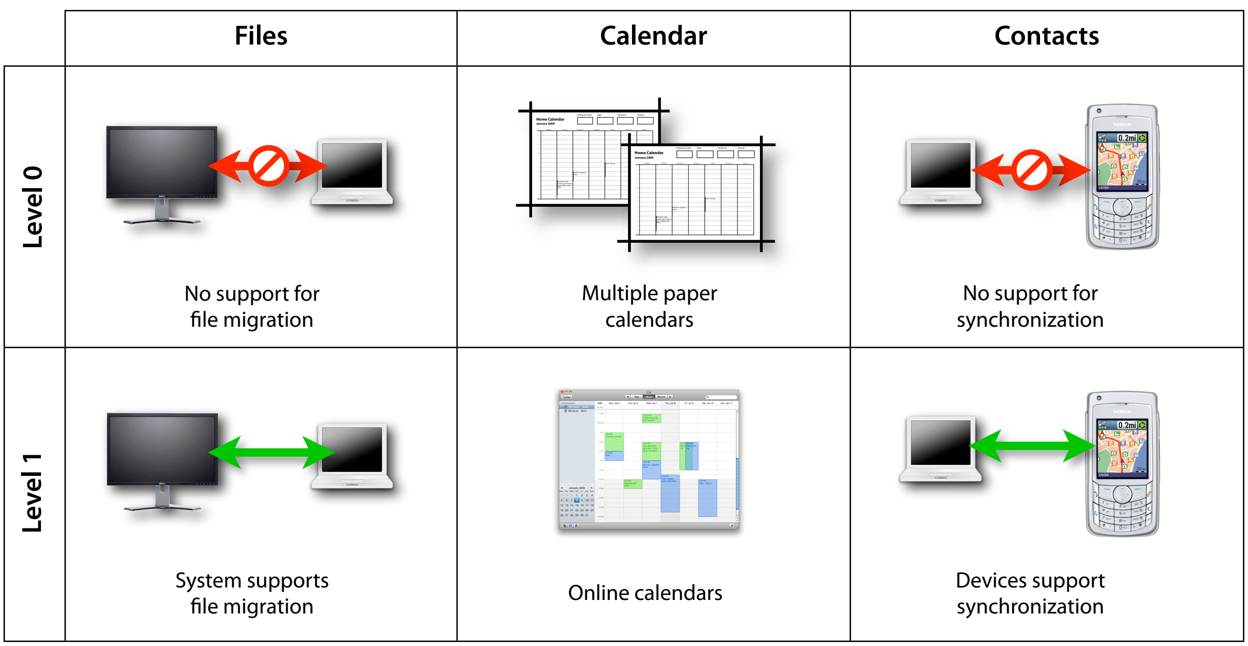


Figure 3.3: An overview of experimental tasks

3.6.1 Pilot Studies

In order to test and validate the experimental setup, pilot studies were conducted with five participants. Feedback from them was invaluable in making sure that the tasks did not take an unreasonable amount of time to complete, and in ensuring that the instructions, steps and other experimental materials were error-free and sufficient to complete the task. Especially in the Calendar task, later instructions depended upon earlier instructions for correctness and completeness; pilot studies provided clues about when this relationship was unclear to participants. Certain instructions were suggestive rather than exhortative, and this led a few pilot participants to misinterpret them. After the experiment was piloted, ambiguous language in a few places was changed to ensure that participants understood the implications of each instruction step they were asked to perform. Data collected from pilot participants was thrown away, and omitted from all statistical and descriptive analyses.

3.6.2 Familiarization Protocol

Since all tasks to be performed during the experiment were common office tasks, it was not deemed necessary to conduct a formal training session, or to require a certain baseline of performance before participants could be recruited. The recruitment criteria used for participant selection ensured that all participants were familiar with the kinds of tasks they were required to perform.

However, training and familiarization can affect the measurements of workload in operators [Eggemeier et al., 1991]. It also was clear that there could be familiarization issues involved if the specific software used was different from what participants were accustomed to. This was especially a concern for the phone task, since different brands of phone have vastly different user interface elements and interaction design. Accordingly, familiarization was provided to participants in two ways, both mandatory.

Demonstration Videos

I created familiarization videos for each piece of software used in the experiment, including a demonstration of the phone provided. Participants agreed (during informal conversations; not statistically significant) that conducting a familiarization procedure with the phone was more important than for desktop software. Experiment participants were not only requested to watch the familiarization videos, but adequate hands-on time was granted to them to gain familiarity with these platforms and use all the devices (desktop, laptop, phone) before starting the experiment.

This familiarization was provided to all participants, regardless of previous experience or use. The videos were available for review, both, before participants arrived at the experiment location, and after they were settled in. Participants were allowed ample time to watch the familiarization videos after arriving for Session 1 as many times as they wished. At the beginning of Session 2, all returning participants were yet again offered the chance to watch the videos. Only a few partook of this offer, however.

Familiarization Tasks

As evidence that familiarization had been successful, each participant was requested to complete 10 familiarization tasks before any experimental tasks were assigned. Familiarization was considered to be complete when participants were able to complete the timed familiarization tasks. A maximum time period of up to one hour was granted to each participant (though none required more than 15 minutes to perform the familiarization tasks). These tasks included simple operations such as editing files of certain types (spreadsheets, presentations), opening the calendar program and creating/editing events, and locating phone numbers and email addresses using desktop software and a phone. The complete set of familiarization tasks is listed in Appendix §7.9.

If participants were unsure about the next steps for any task, or if they explicitly requested assistance from the experimenter, such assistance was readily provided. All questions and queries were handled before proceeding to the experimental tasks.

3.6.3 Subjects and Recruiting

[Eggemeier et al., 1991] stress that workload evaluations be conducted with subjects that are representative of the skill levels expected in operators of the system under study. For this study, I was interested in recruiting knowledge workers who typically use more than one device to perform their personal information management tasks. Students and faculty members at Virginia Tech as well as employees of knowledge-work-related companies at the Virginia Tech Corporate Research Center were thus considered good candidates for this study. This group tends to consist of early adopters of new devices as well as of information management techniques and strategies. Such a population of knowledge workers was relatively easy to locate and recruit on a university campus. Flyers were posted in campus buildings, and email messages were sent to several campus mailing lists.

In return for their time and to encourage genuine effort, participants were compensated with a gift certificate for an unlimited pizza buffet from a local eatery, Backstreets Italian Restaurant[3] in Blacksburg, VA. Since the experiment was conducted in two sessions, participants were eligible for two such gift certificates, and had the freedom to drop out after the first session.

Sample size estimation conducted after 6 participants had performed the experiment revealed that a medium to large effect was evident according to Cohen's d [Cohen, 1988] (details in section §3.6.4). The sample size required to detect such an effect with a power of 0.8 at the $\alpha = 0.05$ level of significance was found to be n=10 for the Files task, n=15 for the Calendars task, and n=15 for

[3]207 S Main St. Blacksburg, VA 24060

the Contacts task. Accordingly, to account for participant mortality, a total of 21 participants were recruited to perform the experiment.

Recruitment Criteria

One of the two screening criteria for participants was that they must be familiar with at least a laptop computer and a cell phone (which all participants met). Other devices such as multi-function phones, laptop computers, and USB drives need not have been used by all participants prior to this experiment, and hence were provided to them before the experiment for familiarization purposes (Section §3.6.2).

The specific model of the eye tracker that was used could not be fitted over eye-glasses worn by a participant. Thus, the second recruitment criterion was that eye-glasses could not be worn. Uncorrected vision and contact lenses were deemed acceptable alternatives.

Since this was a two-session experiment, I allowed for the possibility of experimental mortality (a few participants dropping out of the experiment after the first session), and scheduled a total of 21 participants to perform Session 1. Since this was a within-subjects design, experimental mortality would not lead to dissimilar groups or other unintended side-effects. Out of 21, a total of 3 participants did not attend session 2, and their data was dropped from the final analyses. One participant had to be dropped because of scheduling conflicts for session 2; a second participant was dropped because of data collection issues (related to computer network problems) identified in Session 1. A third participant attended a presentation by me, and thus was made aware of the specific hypotheses and research questions of this experiment. Thus, they were dropped because of a perceived risk of potential experimenter bias.

The study was approved by Virginia Tech's Institutional Review Board under IRB #08-652; the IRB approval form is available as an appendix (§7.3).

3.6.4 Power Analysis and Sample Size Estimation

In order to determine the number of participants required to detect a significant effect, a prospective power analysis was performed, following recommendations in [Lenth, 2001] and [Lerman, 1996]. Since three experimental tasks were performed, power analysis was performed for each one separately. Variance estimates from the first 6 participants were used to compute the sample size required to detect an effect with adequate power. Type I error was controlled at the $p = 0.05$ level of significance for all measures, and statistical power of 0.8 was deemed acceptable for this study (based on both, a review of the statistical analyses reported in similar studies, and guidelines published in the statistical community [Cohen, 1988, Lerman, 1996, Baguley, 2004]).

It is important to note that this was a prospective power analysis, not a retrospective (or post-hoc) power analysis, although the data from the first 6 participants was used for estimating observed power. The results obtained from this analysis are reported solely because they were used for sample size estimation. Specifically, I do not claim that these numbers either express support or lack thereof for any effects observed, statistically significant or not, avoiding the fallacies discussed in [Hoenig and Heisey, 2001]. This is consistent with the recommendations in [Baguley, 2004] and [Lenth, 2001] related to the acceptability of prospective and retrospective power calculations for purposes of sample size estimation and reporting.

Since three experimental tasks were performed, power analysis was performed for each task separately. The NASA TLX scale defines Overall Workload as a weighted sum of the scores along its six individual dimensions; as such, all power calculations were done based on the Overall Workload measure instead of six individual measures. Effect sizes was calculated using Cohen's d [Cohen, 1988] for each task separately. All three were found to show medium to large effect sizes (see table 3.1). Sample sizes were then estimated based on observed effect sizes, α controlled at 0.05, and statistical power = 0.8. The maximum calculated sample size out of all three was used as the effective sample size required.

Task	Cohen's d	Effect Size	Sample Size Estimate
Files	d = 0.67	Between medium and large	$n = 9.78 \approx 10$
Calendar	d = 0.53	Between medium and large	$n = 15.10 \approx 15$
Contacts	d = 0.54	Between medium and large	$n = 14.67 \approx 15$
Across All Tasks	d = 0.60	Between medium and large	$n = 11.86 \approx 12$

Table 3.1: Power analysis calculations for sample size estimation

3.6.5 Experimental Protocol

Upon showing up for the experiment, participants performed the following steps in the order shown. A visual representation of the experimental protocol can be seen in Figure 3.4.

1. **Informed Consent.**

   At the start of the first session of the experiment, participants were requested to take the time to understand and sign a consent form. (See Appendix §7.3.2 for a copy of the IRB-approved consent form used.) The experimenter provided a short background of the experiment and an introduction to some of the tasks they were about to perform.

2. **Demonstration Videos.**

   As detailed in §3.6.2, participants were shown videos on how to use the software in question.

3. **Familiarization Tasks.**

   This was followed by a familiarization session (details in §3.6.2) where users could use the software and devices for as long as they wanted. Most participants interacted with the devices for a few minutes each. They were then requested to perform 6 specific familiarization tasks (details in §7.9.1). These tasks were performed in the same environment and on the same equipment on which the experimental tasks were performed.

4. **Eye Tracker Calibration.**

   Participants wore the eye tracker and performed a 13-point calibration routine twice. A short set of 13 slides was presented to the user. Each slide contained a single white circle about 2 cm in diameter on a black background. As each slide was displayed, one after the other, the participant was requested to fixate on the circle. The experimenter calibrated the eye tracking software in real-time based on the participant's eye fixations.

5. **Experimental Tasks.**

   After the setup and calibration was complete, participants proceeded to performing the experimental tasks (§3.7).

3.6.6 Environment Setup

Participants were provided two computers and one phone. The desktop was an Apple PowerMac G5 and the laptop was an Apple PowerBook G4. Both machines ran Apple Mac OS X Leopard 10.5.6 with all manufacturer-issued software updates applied. For the Files task, they used iWork '09 (Numbers & Keynote)[4], and TextEdit to edit their documents. Dropbox[5], an online storage provider service with an auto-syncing feature was used as the infrastructure for the Network Drive in L1. For the Calendar task (L1 only), they used iCal to manage calendars. For the Contacts task, participants were given a set of contacts pre-populated in Apple Address Book and an AudioVox SM 5600 smart phone running Windows Mobile 2003. Synchronization was performed using a third-party utility, Missing Sync for Windows Mobile[6].

3.6.7 Instructions Display and Time Measurement

As shown in figure 3.5, the desktop was placed to the left of center, while the laptop was placed to the right of center. Between the two, instructions were presented on a large 30-inch display. A custom web application was written to present instructions to the participants, one at a time.

[4] http://www.apple.com/iwork/
[5] http://getdropbox.com/
[6] http://www.markspace.com/

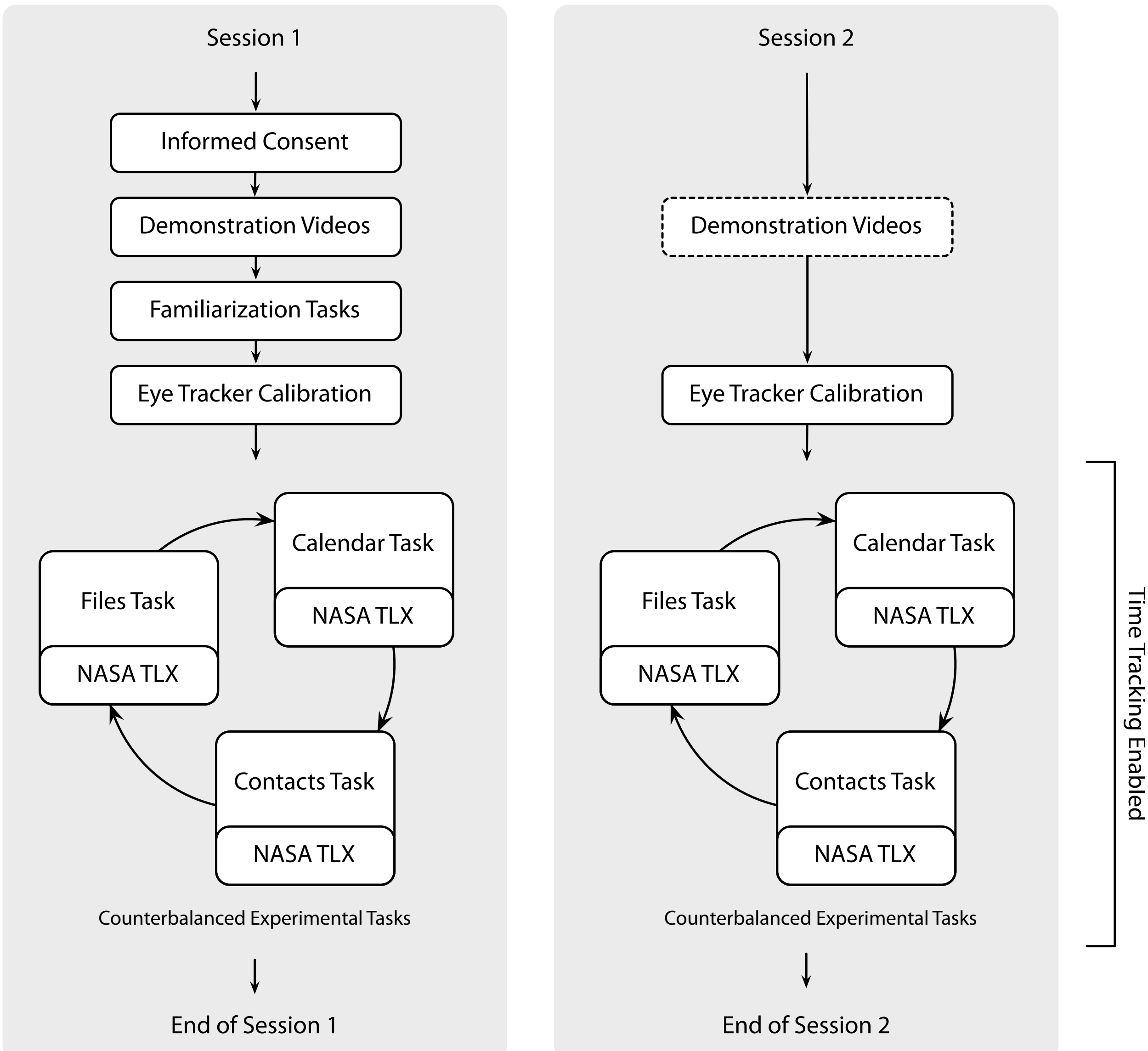


Figure 3.4: Overview of the experimental protocol

When the display changed from one instruction to the next, the app recorded the timestamp. This was later used to analyze sub-task-level changes in physiological measures of mental workload. Table 3.2 shows the readability scores of task instructions, according to two commonly-used readability indices (Flesch-Kincaid and Gunning-Fog.) Figure 3.6 shows the graphical display used for instructions.

A few instructions asked explicit questions of the participant; participants were required to answer these verbally before proceeding to the next step. The experimenter noted down the answers to these questions on paper, taking care not to delay the participant in proceeding to the next step if

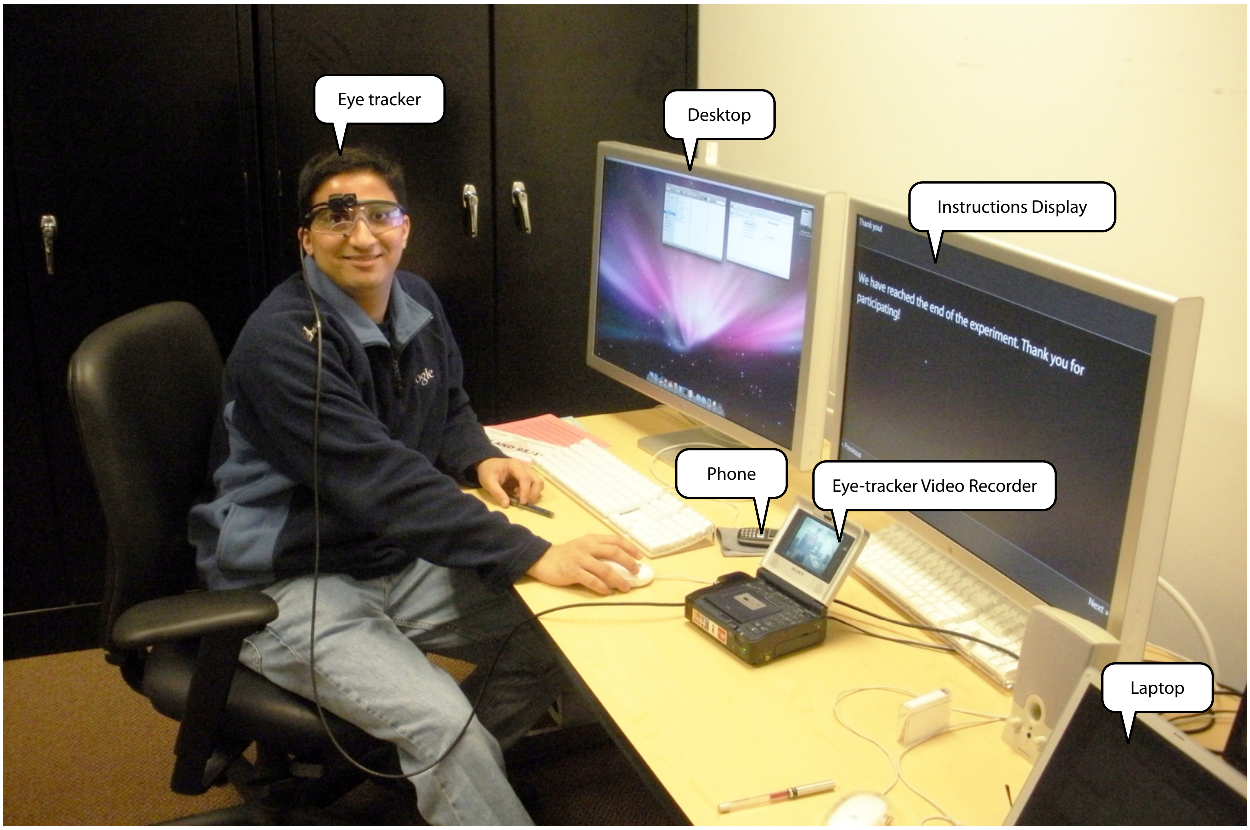


Figure 3.5: Experimental setup

Instructions for Task	Flesch-Kincaid Reading Ease	Flesch-Kincaid Grade Level	Gunning-Fog Score
Files	74.70	6.20	6.90
Calendar	83.00	4.60	6.40
Contacts	78.80	5.20	6.80

Table 3.2: Readability scores for task instructions

she were otherwise ready. This ensured that we were able to capture their responses in the middle of a task, which reflected their true understanding of their personal information (either calendar entries, or contact information) at that point of time. To simulate real calendar entry events, the description used several types of nomenclature: today, tomorrow, Wednesday, Jan 5th, Weekend, etc. It was also interesting to note the specific device that was used for the lookup task. This information would not have been available simply by examining the artifacts (e.g. phone, calendar) post hoc. In a few instances, users realized at a later step that they had answered a previous question incorrectly. But since they had already answered that question, it was possible to capture this mistake through the

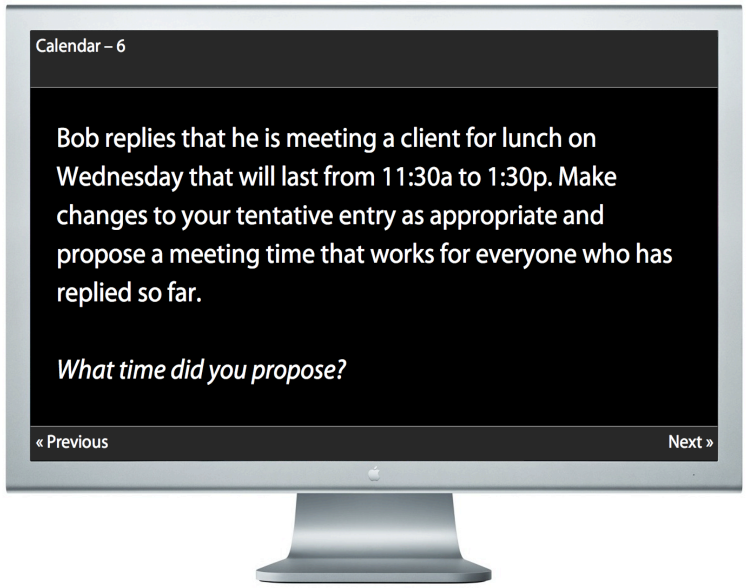


Figure 3.6: Instructions display

experimental apparatus.

3.6.8 NASA TLX Administration

Participants were requested to provide a subjective estimate of workload using the NASA TLX scale after each task. This was administered using a paper-based questionnaire (see copy in appendix 7.7), and participants were asked to place a check-mark on a 20-point scale, as described by [Hart and Staveland, 1988]. This was scaled to the 100-point scale by simple multiplication by a factor of 5. Pairwise comparisons were administered by presenting each pair of dimensions with a checkbox next to each. 15 combinations among 6 dimensions were administered in a random order and participants checked off the option that they deemed more important to the task just performed.

Accurate reporting of workload levels is dependent upon the capability of the operator to recall the experienced workload or effort expenditure associated with performance [Nisbett and Wilson, 1977, Eggemeier and Wilson, 1991] making delayed subjective reports potentially troublesome. Thus, the NASA TLX questionnaire was administered at the end of each task, i.e. once each after the Files task, the Calendar task and the Contacts task, and during both sessions (for a total of 6 such measures per participant over the course of the entire experiment.)

3.6.9 Pupil Radius Measurement

Pupil radius measurement was performed using a mobile head-mounted eye tracker, MobileEye manufactured by Applied Science Laboratories [Applied Science Laboratories, 2007] (Figure 3.7). A desktop-mounted head tracker limits head movement for the participant, which would have caused significant difficulties in this study because of the intrinsic need to interact with multiple devices.

- **Hardware.** The hardware consists of a head-mounted unit that connects via a cable to a video recorder, which connects to a laptop. The head-mounted unit consists of two cameras: one faces forward and captures the scene as viewed by the wearer; the second camera points downwards, and records pupil movement via a transparent mirror that is partially reflective in the near-IR and IR ranges. The laptop runs custom software that allows experimenters to start and stop recording, calibrate participants' eye gaze to specific points visible in the scene, capture pupil data as a comma-separated value (CSV) file, and eye-gaze data as a video file. The eye-gaze video consists of the scene superimposed with a cross-hair at the position of the eye fixation as calculated by the software. The eye tracker records pupil radius at 30 Hz. Details of this procedure are available in the manufacturer's manual [Applied Science Laboratories, 2007].

Figure 3.7: ASL MobileEye eye tracker. (photo by Manas Tungare)

- **Illumination.**

  Illumination was carefully controlled to be the same for all participants and at all times of the day. The experiment was conducted in a closed room, and no external light was allowed to enter the room. When moving between devices, participants moved one meter away from their previous position, with the same orientation, to minimize any potential changes in illumination. All participants sat at the same distance from the display, about 60 cm.

- **Post-processing Pupil Data.**

  A significant amount of post-processing of pupil radius data was needed. The start time of the experimental tasks was synchronized with the start time of the pupil data recorded. Task times were automatically recorded by the instructions display application, described in section §3.6.7. Data for each session was split into separate measurements for each of the three tasks, based on the time at which the participant completed one task and moved to the next. All instances in the time series when the pupil data could not be captured by the eye tracker (due either to blinks, or because the participant was looking at an angle too far outside the range in which pupil radius could be computed) were discarded.

- **Signal Smoothing.**

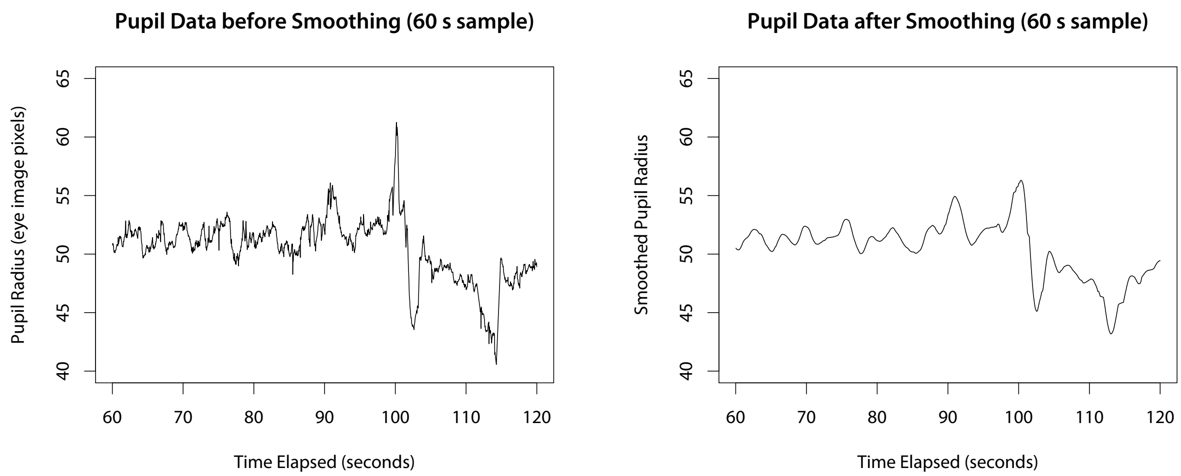
  The raw pupil data was extremely noisy and needed to be smoothened to isolate the signal from the noise. I considered several moving average algorithms for signal smoothing, and finally settled on using the Savitzky-Golay filter [Savitzky and Golay, 1964]. This filter is considered better than a simple moving average filter because the weighted polynomial fit applied over $2n+1$ points tends to preserve distinctive features of the signal while still removing noise discriminatively. For pupil data, I applied a 4^{th} order Savitzky-Golay filter of length 151. I re-used code written in R by Borchers [Borchers, 2004] for this purpose. Figure 3.8(a) shows the raw pupil data collected during 1 minute of activity; figure 3.8(b) shows the same data after running the Savitzky-Golay filter.

- **Baseline Adjustment.**

  After smoothing, pupil radius data was adjusted to account for individual differences in pupil size. A baseline reading for pupil radius was obtained for each participant from the first 5 seconds of pupil activity data. During the first five seconds, participants were not assigned any specific task or provided any instructions to read, and was considered a period of minimal task-induced workload. Observed pupil radius measurements were scaled by the baseline reading, following a procedure similar to the one reported in [Iqbal et al., 2005].

- **Obtaining Per-Instruction Estimates of Workload.**

  While changes in pupil data were visible continuously as participants performed tasks, I needed estimates of workload per instruction in order to be able to compare the two treatment levels. To obtain a per-instruction estimate of workload, I calculated a simple mean of all pupil data measurements taken during that step.



(a) Raw Pupil Data (b) Smoothened Pupil Data

3.7 Experimental Tasks

When choosing experimental tasks and instructions, I picked those that exhibited the desirable properties of reference tasks for Personal Information Management, according to Whittaker. [Whittaker et al., 2000]. Thus, these tasks were chosen to be those that are *frequent*, *critical* and *real*. In each subsection below is a short description of each of the three tasks at two levels, devices that were used, workarounds that were suggested by survey participants, and the specific questions of interest in each task (above and beyond the general ones stated earlier).

3.7.1 Task 1: Managing Files on Multiple Devices

Over the last few years, users have begun using multiple devices in addition to their own personal machines for nomadic work. A common use case is when they appropriate and use a semi-public computer for their temporary information processing needs and intend to resume processing on their personally-owned machines soon afterwards: e.g. at a library, during laboratory sessions at school, or at the office. Another common use case is when users own two computers, typically a laptop and a desktop, and need to access and modify files on both devices depending upon the context of use. During these scenarios, a common tool for one-time (not repeated) migration of files and documents is the USB key drive, a simple device that may be used to manually copy and move files from one machine to another. Although we did not specifically list this as a computational device, several users pointed out to us that they made extensive use of USB drives in their information management workflows.

When transporting files for such a purpose, there are two related goals: the first goal is to transfer the data from one device to another, without concern for its location. Thus, copying files to USB drives, sending email to oneself with the file attached to it, as well as using network storage, all fulfill this goal. The second goal is to be able to place a transferred file in its correct location on the local disk. This goal depends on the success of the first goal, but does not automatically follow from it. From the examples listed above, only network storage fulfills the second goal; USB drives and email do not assist the user in the placement of a transferred file in its logically correct location.

Although synchronizing data automatically via software is an option, many users chose not to use it because of its complexity. Among those survey respondents who talked about synchronizing data (n=45), the most popular devices were laptops (n=21), followed by phones (n=14), desktops (n=13) and PDAs (n=12). The most common problem was that synchronization failed (n=10), or that their software would not let them sync their data in the way they wanted (n=8). Another common complaint was that data was deleted by the system when the user did not expect it (n=7). This included cases where the sync software overwrote fresh data with stale data from another device, and any other situation that was not directly caused by user error (such as setting the wrong parameters when synchronizing).

Task Scenario

Participants were placed in a scenario where they played the role of a consultant who worked with several clients, either at their own office on the desktop computer, or at one of the clients' sites, using their laptop computer. Participants were asked to identify which of three different file hierarchies appeared most similar to the way they organized their own files (Figure 3.8). Depending on their choice, they were provided with a file hierarchy that was most similar to their own. A set of scripts was used to create either of the three hierarchies at the participants' request.

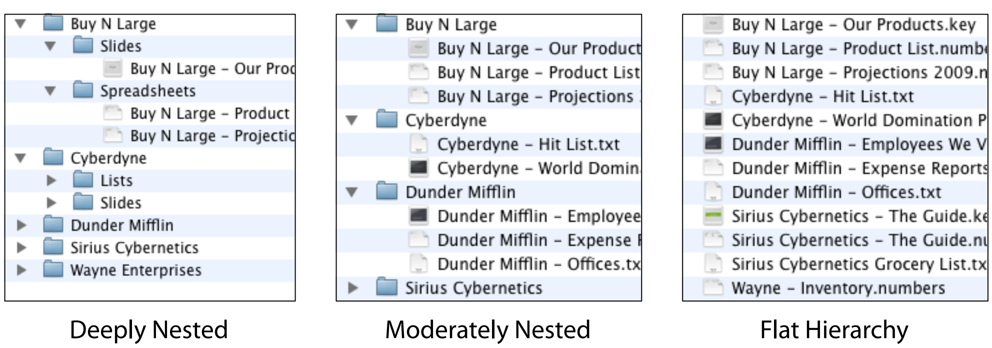


Figure 3.8: File hierarchies: Deeply Nested, Moderately Nested, and Flat

- **Deeply Nested.**
  One directory was created for each client, and within it, one directory for each project.

- **Moderately Nested.**
  Individual directories were created per client, but the files related to all the projects of that client were placed in the same directory.

- **Flat.**
  All files were pooled in the same directory, and enough information was added to the filename for users to be able to infer the specific client and project to which that file belonged.

An exact replica of the chosen file hierarchy was made available on a laptop computer at the start of the experiment. Participants were provided instructions, one at a time, asking them to make certain specific edits to files. There was no ambiguity in identifying the file or the client from the text of each instruction. Files were of different types: presentations, spreadsheets, and text files. A complete list of instructions presented to the participants is available in appendix §7.9.2.

After a few such edits were made, participants received instructions that they now had to wrap up their work at their own office and travel to a client site. Before they did so, they were asked to take whatever actions they felt necessary to be able to have access to their files from the client site.

In L0, they were provided USB drives for transferring their files back and forth. They were also provided access to web-based email with an account they could access from either machine. In L1, software support was provided for remote access to their files via a Network Drive.

After the physical move from the desktop to the laptop, participants were allowed time to settle down and take any actions at their discretion before they declared themselves to be ready for fresh instructions. During this time, they either copied files from USB drives back to the laptop, accessed network storage, or downloaded the files sent to their own account via email. Multiple instructions followed this phase, and finally, participants were requested to make a reverse move, back to their desktops. They were allowed time to ensure that their documents could later be accessed from the desktop. The task was deemed complete when the participant announced that they believed they have completed the transfer of files from the laptop to the desktop.

If participants asked questions related to where the Pictures folder was, or were not sure where the final documents should be copied to, those questions were answered by the experimenter. Other questions such as "should I move the files or copy them?" were answered with a non-committal response intended to capture their own methods: "please do as you would do with your own files/-calendar/contacts".

Treatment Levels

- **Level 0:** Provide users a USB drive and an email account to transfer files back and forth.
- **Level 1:** Provide users a Network Drive to transfer files between their own machine and the device they have recently appropriated for use.

Devices

For this task, users were provided a desktop computer and a laptop computer, with specific files placed on both machines in a specific location. In L0, migration solutions such as USB drives and email accounts were provided, while in L1, a network storage location accessible from both computers was provided.

Issues of Interest

We wish to examine whether the mental workload caused is the same or different in two cases where users are required to move their files between devices.

Measurements

After each experiment, several artifacts were collected and analyzed. Files saved to the desktop and laptop, those saved to the USB drive, or Network Drive or email account were all archived. Timing was measured by a custom web application that was used to administer instructions to each participant. This web application is discussed in detail in section §3.6.7.

The following task-dependent performance metrics were measured from the gathered artifacts:

- Time taken to complete the entire task, and each sub-task (step);
- Time taken to move between machines;
- Number of files correctly edited;
- Number of files placed in their expected location after moving;
- Number of files copied to the USB drive or Network;
- Number of files copied to the Laptop;

3.7.2 Task 2: Accessing and Managing Calendars

Calendar management involves two major goals for users: (1) being aware of events when they are scheduled, and (2) being able to answer questions about their schedule when needed, in a format suitable for consumption depending on the current context and devices. Often, users are interested in events scheduled by others on their own calendar or on a shared calendar. Being able to schedule events on one's calendar from external sources is an important PIM task [Tungare and Pérez-Quiñones, 2008a].

Task

At the start of the calendar task, users were provided either (1) two paper calendars labeled 'Home' and 'Work' (L0) or (2) an online calendar program with two overlapping calendars in it, also labeled

'Home' and 'Work' (L1). During the task, participants were presented instructions that required them to consult or update their calendars. Different types of events were included:

- **Tentative events:** An event was present on the calendar, but marked tentative. The event description also included a description of how the event could be confirmed ("call your spouse to confirm this event").
- **Rescheduled events:** An event on the calendar was moved to another day, but at the same time. Of interest was the question whether any differences are seen in such an operation between paper and online calendars.
- **Group events:** Group events involve multiple communications among attendees to settle on a time that works for everyone. This situation was simulated in the experiment by a series of instructions that provided one piece of information at a time. These instructions were not provided as a single uninterrupted sequence, but were interleaved among other calendar instructions.
- **Events that require preparation:** Several events in real life require preparatory actions to be performed, e.g. driving to an event. This aspect was included in the experiment by explicitly mentioning the driving time required to attend an event. The participants were then asked to schedule the event for whatever chunk of time they saw fit (i.e., they were not specifically instructed either to include or exclude driving time.)
- **Conflicting events:** Participants were asked to express their availability for events that were clearly conflicting. In addition, one tentative event was set to conflict with a meeting request such that the correct answer about availability was neither yes nor no, but "I need to check with my spouse about this other tentative event."
- **Pre-planned and unplanned activities:** Some events were planned up to several days in advance, whereas some events were scheduled (or were asked to be scheduled) only a few hours in advance.
- **Queries about free time:** Questions asked in the instructions were not limited to scheduled events only; a few instructions specifically queried the participants about their free time ("When on Friday can you go to the dentist?")

The calendars that were provided to them already contained several meetings scheduled (confirmed or tentative, marked clearly), so that the events presented during the experiment were not sufficient in and of themselves to answer the questions posed. In other words, this forced them to consult the calendar instead of simply memorizing the information presented earlier.

Different types of calendars were used in both sessions.

- **Level 0.** Multiple paper calendars were made available to the user, and they were instructed to add newly-scheduled events to aid future recall. One calendar contained 'Work' events,

while the other contained 'Home' events. A few overlapping events were scheduled on both calendars.
- **Level 1.** A computer-based calendar program was provided to participants, and individual calendars were assigned specific colors (a common feature of calendar programs). The default view afforded viewing multiple calendars overlapped together.

Devices

The paper calendar session involved no devices, while the online calendar program was administered on a desktop computer.

Issues of Interest

Understanding whether interpreting overlapped events in a calendar results in lower or higher mental workload than consulting multiple individual calendars when making scheduling decisions.

Measurements

After each experiment, artifacts of calendar use were collected and analyzed: these include copies of the paper calendars and screenshots of the online calendar program. Timing was measured using the same web application as in the Files task.

The following measures of task performance were used:

- Number of calendar entries made successfully;
- Time taken to complete each step;
- Errors in responses regarding one's schedule.

3.7.3 Task 3: Managing Contacts Using a Phone

A majority of problems encountered by our participants regarding phones involved software limitations or interface deficiencies. The presence or lack of synchronization options was seen to affect users' information management strategies. In L0, participants were expected to perform contact management tasks using a laptop and a phone with no synchronization software support. In L1, they performed the same tasks using the same hardware, only with synchronization support provided by the system.

Task

The most common task performed on a cell phone, after making phone calls, is locating contact information. A particularly frustrating aspect that was highlighted in the survey involved navigating the address book on a cell phone, and syncing information between their computer and their phone. Accordingly, the experiment required them to recall the phone numbers of people whose phone numbers existed solely on their laptop address book.

Participants were described a scenario where they were a researcher attending a conference, and met several old acquaintances and made new contacts. They were allowed to access their laptop at some times, and their phone at other times, and both at some other times. When attending a paper presentation session, they could use their laptop, but not their phone; in the hallways, they could use their phone, but not their laptop. At the end of the day in their hotel room, they were free to use whichever device they preferred.

Instructions specified clearly whether or not they had access to their laptop and/or phone at each point of time. It was also required to enter the information on the device specified in the instructions. We refer to this device (either a laptop or a phone) as the **primary device** in the forthcoming discussion. The other device (either a phone or a laptop), to which they did not have access per the instructions, is termed the **secondary device**. The distinction between the two is not by any specific role of the device, but solely based on what the instruction specified. Thus, obviously, both devices were primary or secondary at different points of time, depending on the instruction on screen.

- **Level 0.** Manage contacts using a laptop and a phone, but without synchronization between the two.
- **Level 1.** Manage contacts using a laptop and a phone with synchronization software available.

Devices

A cell phone and a laptop were provided in both sessions. Synchronization software was provided in L1.

Issues of Interest

Participants have indicated that navigating interfaces, especially on cell phones, can often be painful (n=18). Navigating the address book on a phone is especially troublesome, and so is trying to sync a phone with a laptop.

Measurements

As with the first two tasks, all products of interaction were collected after each experiment: this included a complete list of all contact records from the desktop address book, as well as contacts collected from the phone address book. Questions were asked to participants that involved them looking up either the email address or phone numbers of certain persons.

- Time taken to enter or lookup contact records;
- Number of contact records created or edited on the primary device;
- Number of contact records created or edited on the secondary device;
- Number of errors in answers regarding their schedule and availability.

3.7.4 Constraints & Limitations

PIM Experiments are Impersonal

One of the most common challenges and limitations in Personal Information Management studies is that experimental tasks can never be as personal as a user's own information collections [Kelly, 2006] — only a close enough approximation. Even minor changes in style and layout can cause subtle changes in user behavior: e.g. one of our participants reported that their own calendar program is setup to display dates from Sunday—Saturday for any particular week, while the calendar program provided during the experiment was set up for Monday—Sunday.

Any kind of experimental study in PIM suffers from this limitation. The natural environment of a user's information ecosystem cannot be recreated in a laboratory setting. While ethnographic and other field work approaches can provide rich descriptive analyses of a user's practices, they could not have yielded the data we collected in this experiment.

Despite this, we took every care to pick tasks that would be very similar to what knowledge workers would be exposed to during their regular lives. Familiarization was provided so they would be able to examine the software and understand the information already entered (e.g. files, calendar entries, contact records, etc.) To bring the experimental setup closer to users' personal idiosyncrasies, we provided three alternate file system hierarchies to pick from. The details are provided in section §3.7.1.

Generalizability

Since all participants were knowledge workers and had a minimum college-level education, it cannot be predicted whether these results would be generalizable to the rest of the population. While this is a limitation of our sample, such a population is representative of the users who perform PIM tasks in their professional lives, i.e. knowledge workers.

3.8 Analysis of Study 2: Statistical Tests

All statistical computations were performed using the R Language and Environment for Statistical Computing and Graphics [R Development Core Team, 2008]. Power analysis, in particular, was done using the `pwr` package [Champely et al., 2007] which is based upon Cohen's effect size calculations [Cohen, 1988]. Savitzky-Golay filtering applied to the pupil radius data was performed using code written by Borchers [Borchers, 2004]. All scripts used for cleaning up and analyzing the data are included in an appendix to this dissertation (§7.10).

A note about the applicability of metrics used in the analysis: wherever NASA TLX scores are analyzed, they are for the entire task. Wherever pupillometric estimates of mental workload are analyzed, it will be made explicit whether the estimate applies to the entire task, or to each step of the task (i.e., instruction) individually. Several steps in each task are performed similarly in the two levels (L0 and L1), and differences in pupil radius cannot help discriminate between these specific steps.

In this section, I describe the specific analyses I conducted to test the hypotheses presented in §1.3; results are presented in the next chapter (chapter 4).

3.9 Testing Hypothesis H1

Restating Hypothesis 1 from §1.3.1: I hypothesize that the variability in workload imposed by dissimilar tasks will be high. The level of support provided by the system for task migration affects mental workload: higher level of support would lead to lower levels of workload and vice-versa.

To test this hypothesis, I conducted two analyses.

3.9.1 NASA TLX Scores across Tasks and Treatments

I conducted seven 2-factor Analyses of Variance (ANOVA) of NASA TLX scores across Tasks and Treatments, one for each dimension of the NASA TLX scale, including Overall Workload. The R script used to perform this analysis is available in appendix §7.10.5.

3.9.2 Task-Evoked Pupillary Response across Tasks and Treatments

NASA TLX provides workload estimates at a task level, while pupillometric data provides a continuous workload estimate. I conducted an analysis of variance of the adjusted percent change in pupil radius across treatments for each step. Thus, step I1 of the Files task at L0 was compared with step I1 of the same task at L1, etc. The R script used to perform this analysis is available in appendix §7.10.7.

Results of testing this hypothesis are available in §4.3.

3.10 Testing Hypothesis H2

Restating Hypothesis 2 from §1.3.2: I hypothesize that operator performance measured via each of these metrics will be higher when there is a higher level of system support for task migration.

Task performance was obtained for each task separately, using task-specific metrics. The only measure of task performance that was used with all tasks was time on task. Other task-specific metrics are discussed in sections §3.7.1, §3.7.2 and §3.7.3. As stated earlier (§3.8), time-on-task measures were available per task as well as per instruction. Accordingly, I performed two analyses to test this hypothesis.

3.10.1 NASA TLX Scores and Task Performance (per Task)

I attempted to correlated NASA TLX scores for each Task and Treatment level with performance metrics measured for each task.

3.10.2 Task-Evoked Pupillary Response and Task Performance (per Instruction)

In every task, there are a few critical instructions that contribute to sudden changes in workload. In order to identify and highlight such cases, I attempted to correlate pupillometric estimates of workload (corresponding to the task-evoked pupillary response) with per-instruction performance metrics.

Results of testing this hypothesis are available in §4.4.

3.11 Testing Hypothesis H3

Restating Hypothesis 3 from §1.3.3: Subjective measures and physiological measures of workload will correlate with task performance metrics and with each other during the execution of a specific task.

Since subjective measures of mental workload are only available at the end of each task, the task-evoked pupillary response used for testing this hypothesis was also coalesced to the mean of all readings obtained during the performance of a specific task. I attempted to correlate NASA TLX scores against the adjusted percent change in pupil radius, measured over the course of the entire task.

Results of testing this hypothesis are available in §4.5.

Chapter 4

Results

As described in chapter 3 in section §3.2, the first study was a survey to understand current practices in multi-device PIM and to develop representative tasks for the second study. In this chapter, I present the results from the two studies and provide a short description of the results. The implications and discussion of the results are included in the following chapter.

4.1 Results from Study 1 (Survey)

The first study was strictly exploratory: the goal was to collect information from a wide base of users about the devices they used, the activities they performed, the problems they faced, their solutions, etc. The knowledge gained from this study was used to inform the design of the second study (controlled experiment.)

4.1.1 Participant Demographics

220 respondents completed the survey. Since a link to the survey was posted to several message boards, mailing lists and web sites, an accurate count of the people it reached could not be estimated; thus, the survey response rate is not available. 53% of respondents were male, 30% were female, and 17% indicated neither. 157 respondents, or 71.3%, reported that they considered themselves either full-time or part-time knowledge workers. The study spanned an age range from 18 years to over 58 years old. Though a majority of the respondents were between 22 and 30 years old, other age groups were adequately represented (see Figure 4.1).

The participant pool consisted of users of varying levels of education completed, from high school to doctoral degrees. Due to our focus on knowledge workers, our study elicited a high number of responses from people who had completed advanced graduate degrees: Masters 34% and Ph.D. 10%.

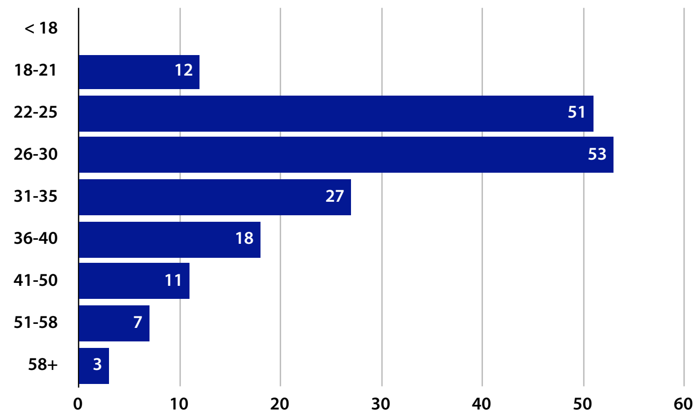


Figure 4.1: Number of survey respondents by age group.

4.1.2 Devices Used

Figure 4.2 shows the number of each type of device reported, converted to percentages. Our study found more laptop users than desktop users. Over 71% of respondents used at least one desktop, while about 96% used at least one laptop, which is higher than even the number of cell phones reported. This is representative of the current trend towards mobility and away from stationary platforms such as desktops. Portable media players have made their way into the hands of more than 80%, almost equal to that of digital cameras.

Handheld computing devices that combine a Personal Digital Assistant (PDA) and a cell phone, such as the Blackberry, Palm Treo, Apple iPhone and others, are used by a minority of users, about 22%. PDAs without built-in cell phone technology are used by fewer users, about 15%. These results must be interpreted within the context of the time at which this study was conducted (August 2007). The distribution of these devices is likely to be different at the time this dissertation is published (March 2009) due to various market conditions affecting the sale of such devices.

4.1.3 The Impact of Multi-Function Devices

Those who use multi-function devices use them extensively for PIM tasks such as email, calendaring and instant messaging (IM), and also for news and (limited) Web browsing. Some participants reported that the presence of these handhelds had caused them to leave their laptops behind when they did not expect to work on complex documents (e.g. when on vacation), but many others reported that they still carried their laptops with them as the tool of choice for more complex computing activities.

> *"Treo allowed me to stop carrying a separate pager. I still carry a laptop around. How-*

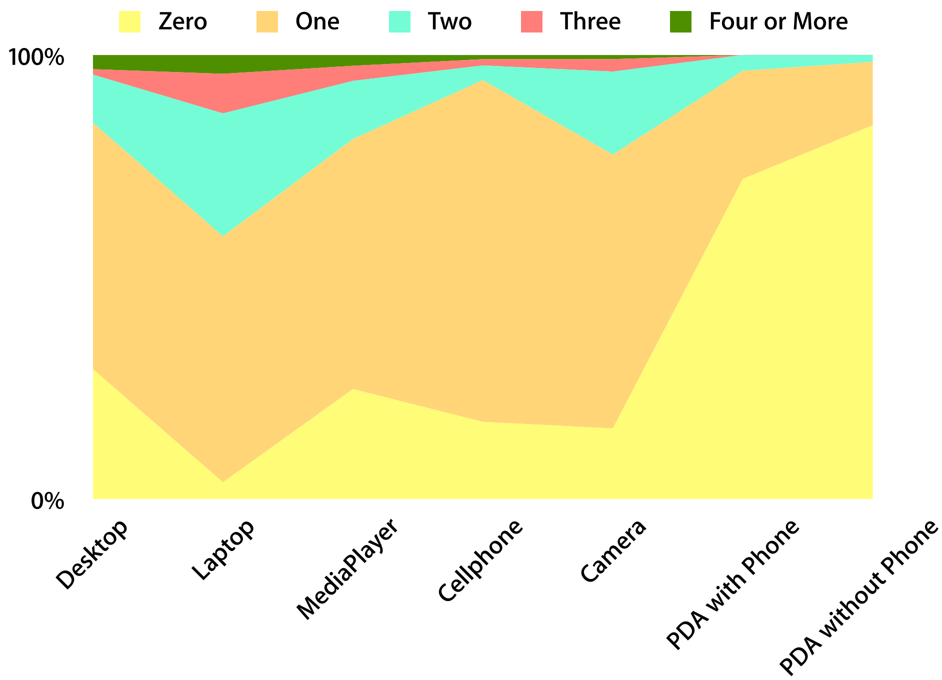


Figure 4.2: Number of devices in each category, reported as percentages.

ever, when I don't have the laptop, I can still do almost everything – except edit documents – on my Treo."

These multi-function devices often replaced other single-function devices, like cell phones, music players, and compact digital cameras. Participants reported that they started using more features of their device because they carried it with them for another purpose. Certain activities also were moved from one device to another simply because it was now possible to do so, without the burden of carrying yet another device. This is an example of an unforeseen (or unintentional) advantage of acquiring a new device. The quality of individual functional components of an integrated device was often compared to that of stand-alone devices and generally found to be lacking. Despite that, the convenience of carrying a smaller device led users to prefer them on certain occasions. Features of devices that did not integrate well within their existing information infrastructure were used less often. Users reported that synchronizing with their other devices was an important requirement, irrespective of the quality of the stand-alone feature.

"I have a Windows Mobile Smartphone with a full keyboard. [...] Its camera isn't near as good enough to replace my digital camera and the calendar doesn't sync with my MacBook, so I don't use it."

4.1.4 Groups of Devices

Several users reported that they used devices together in groups. Figure 4.3 shows the most common device groups. (Device groups reported by fewer than 10 participants are not included in the figure, to avoid cluttering and hindering insights.) Laptops and cell phones were used together by the most number of users, almost 24%. The laptop and the cell phone also appeared the most times in combination with other devices. The low use of PDAs without an integrated cell phone for almost all tasks (as compared to the use of cell phones and PDAs with cell phones) indicates that these devices are considered less popular.

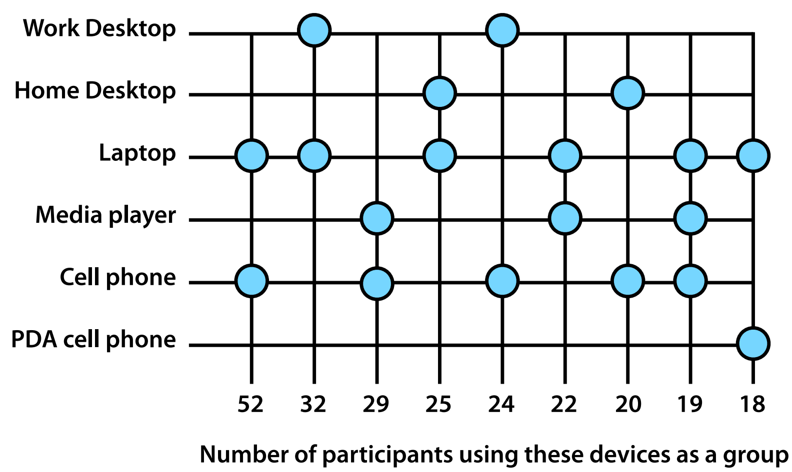


Figure 4.3: Devices used in groups as indicated by survey participants.

However, a lot of users were dissatisfied with the currently available synchronization tools for multiple devices. The high use of the laptop is indicative of the trend to keep all data on a single device to escape the need for synchronizing. Similarly, address books on cell phones were kept separate from those on laptops (or desktops). The same data (or application) was used for two distinct tasks (sending email from the laptop versus making a phone call using a cell phone), and therefore some users preferred to keep the two contact databases separate, again a compromise.

> *"Usually my contacts on the phone are just with numbers while my contacts on the computer are just with email addresses (makes sense since I'm using the former to make calls and the later to send emails). [...] The name of the contact is usually different for emails (e.g. full name instead of only first name or last name first or use of title in front of name.)"*

4.1.5 Activities Performed

Given the vast array of devices and the features they each support, we wanted to learn what features actually are used. The laptop and desktop were reported as the ones where most computing activities were performed (see Figure 4.4; the diameter of each circle is a logarithmic function of the number of participants who perform a given activity on a given device.). Mobile devices such as cell phones and PDAs were used for contact management, making phone calls, and calendaring, and to a lesser extent for browsing and instant messaging. None of the users viewed or edited documents from devices with a smaller form factor than the desktop/laptop. Users had trouble browsing through their data on small devices, and reported skipped adding more data in order not to "pollute" the pool of data already on the device.

I found several instances of activities performed across devices: users tethered their laptop to their cell phone so that the cell phone's network connection could be used by the laptop, without having to forfeit the richer form factor of the latter. Music was moved off the laptop onto the media player because the media player always was at hand in addition to the laptop.

> *"I typically will take down someone's email or phone number on a sticky note and then affix it to my cell phone. I find my cell phone's contact navigation to be a real pain. Thus I find it tedious/somewhat-pointless to put more people on there – after all it will just cause me more pain when I am navigating to people I really want to call."*

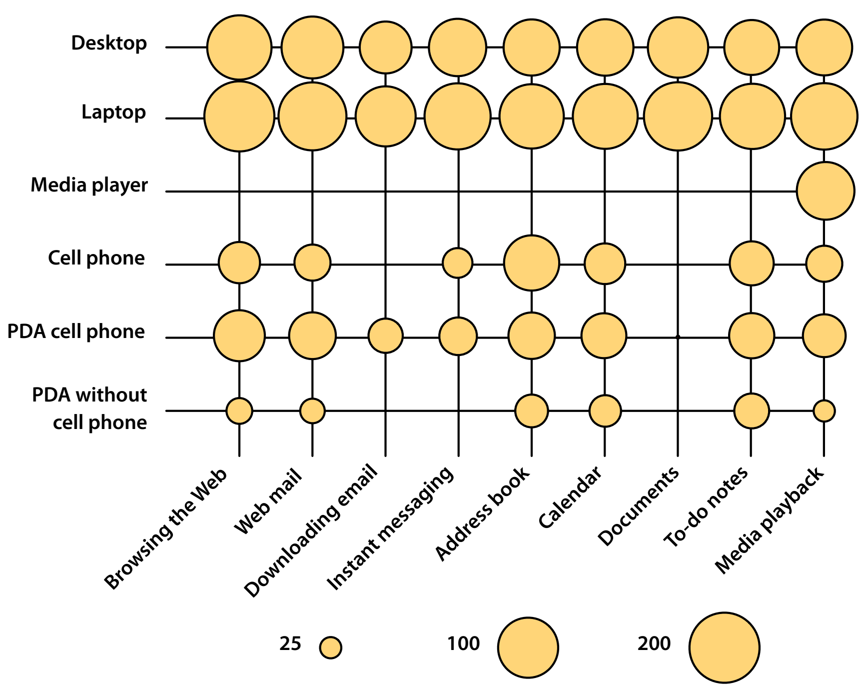


Figure 4.4: Activities performed by users on devices.

4.1.6 Content Analysis of Qualitative Responses

The following table illustrates the results of the content analysis of data collected in Study 1. Table 4.1 contains a complete list of the values for each tag that were determined during the analysis. In the content analysis I performed, the categories of tags were determined *a priori*, while the actual tags were determined on an emergent basis. The following figures show a list of the devices users reported (Figure 4.6), the problems they encountered (Figure 4.5) while trying to perform their tasks (Figure 4.7), the solutions they came up with (Figure 4.8(a)) and the final outcome (Figure 4.8(b)), whether or not they were successful.

Tag	Domain Values
`device`	`laptop, desktop, pda, phone, mediaPlayer, multiFunction, camera, server, usbDrive, externalDisk`
`task`	`browseTheWeb, downloadEmail, accessWebEmail, instantMessaging, sendTextMessages, makePhoneCalls, accessContacts, accessCalendar, accessDocuments, toDoLists, takePhotos, playMusic, playVideo, syncData, doBackup, setupNetworking, setReminders`
`problem` (System)	`backupFailed, cannotAccessData, collaborationDisallowed, conflictingEdits, conflictingVersions, duplicateData, formatIncompatibility, hardwareFailure, hardwareIncompatibility, interfaceDeficiencies, lowFidelityData, metaDataIncorrect, networkingNotWorking, noSoftwareExists, policyRestrictionsOrDRM, setupAndInstallation, softwareLimitations, syncFailed, syncParametersIncorrect, systemCrashed, tooHeavy, tooManyDevices, tooManyFeatures, unexpectedDeletion`
`problem` (User)	`forgotToSync, notBackedUp, deviceStolen, accidentalDeletion, noConnectivity`
`solution`	`copyManually, emailToSelf, everythingOnline, maintainSeparateCopies, printCopy, replicatedStorage, transcodeFormats, usbDrive, useBackup,useSingleDevice`
`result`	`workedButNotEasy, liveWithIt, workedAsExpected, workAroundSuccessful, workAroundFailed, taskIncomplete`

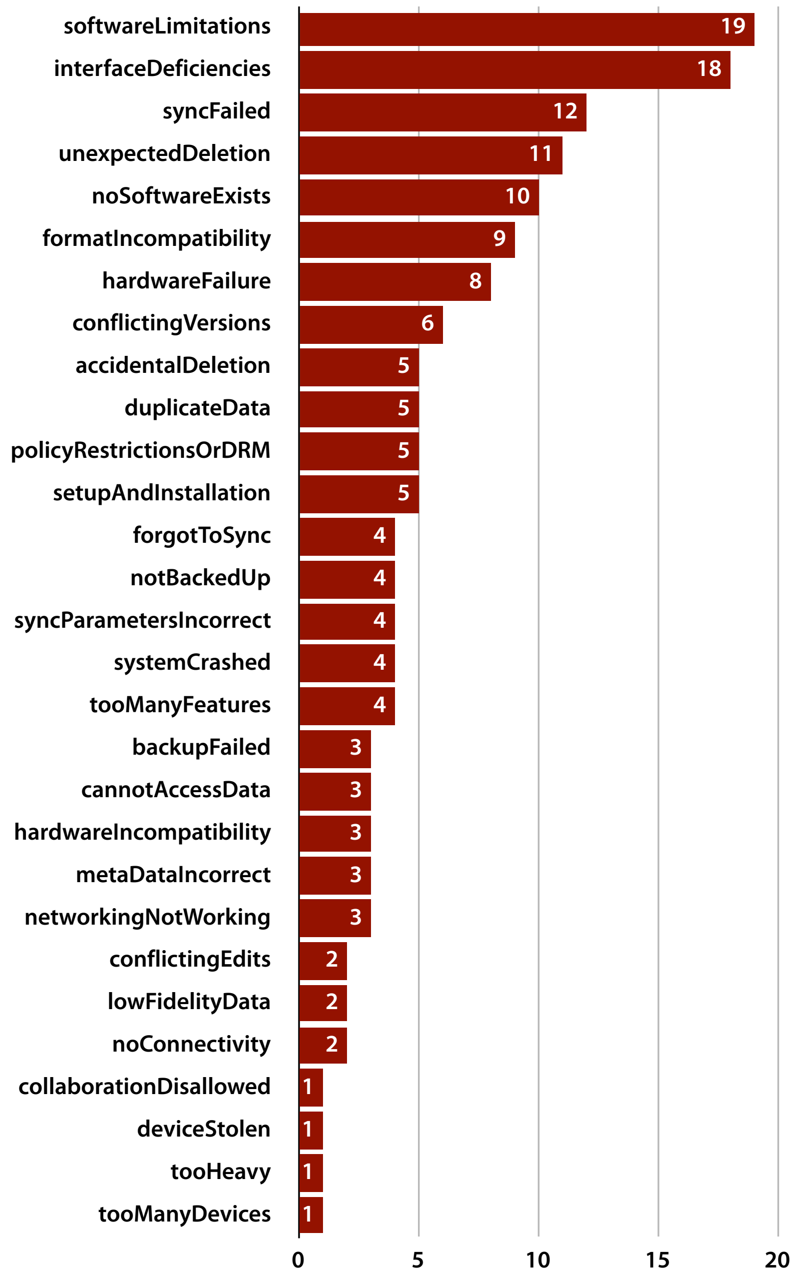


Figure 4.5: Problems that users encountered while completing their tasks.

4.1.7 Commonly-Reported Problems

In the qualitative portion of the survey, users reported several problems that they encountered often when managing personal information across multiple devices. Figure 4.5 shows a list of problems

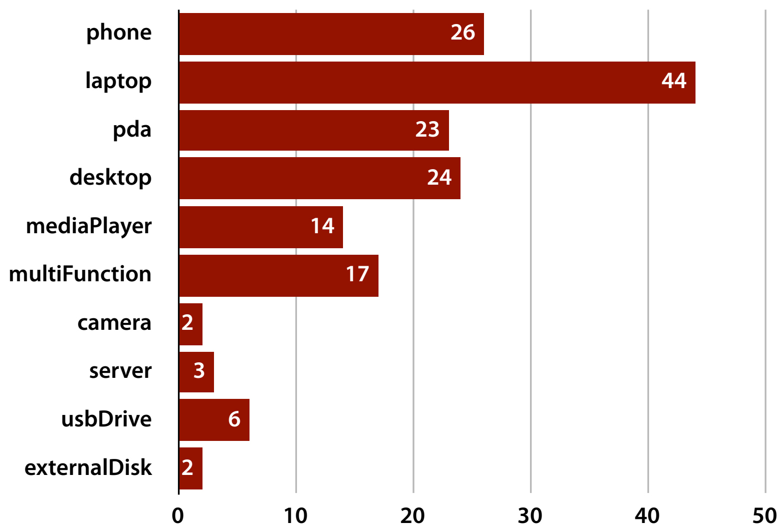


Figure 4.6: Devices reported by users in the questions about their problems.

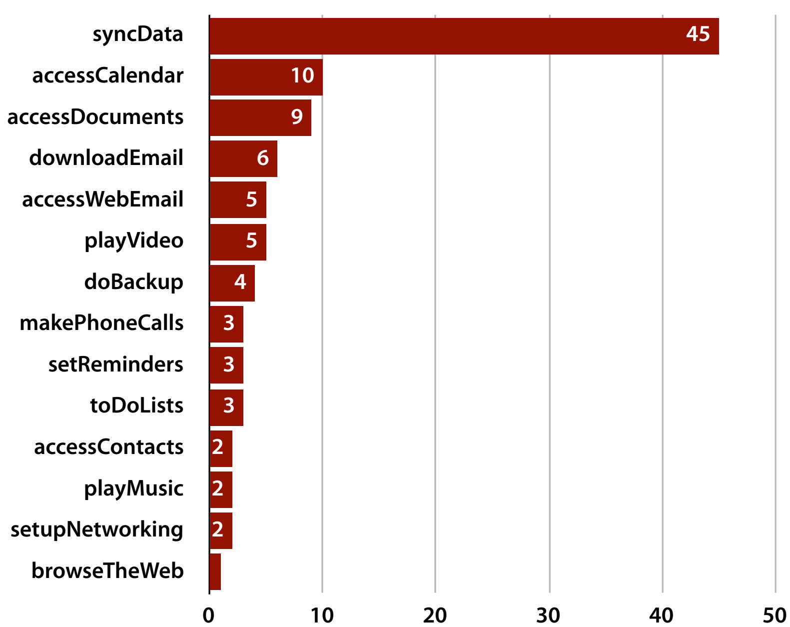
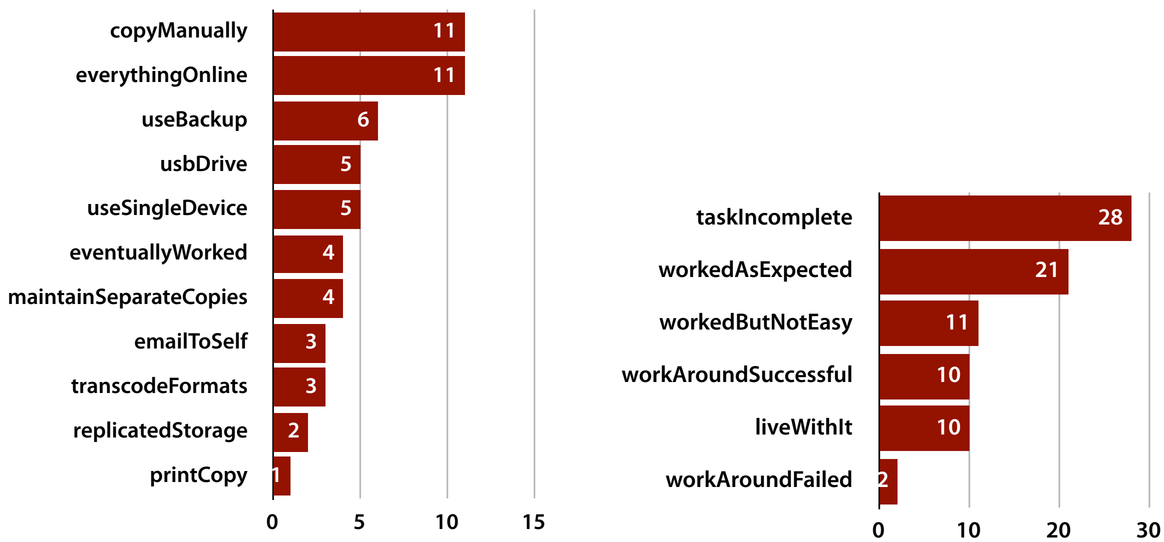


Figure 4.7: Tasks that users were trying to perform.



(a) Solutions developed by the users to overcome the problems.

(b) Resulting outcomes for users.

Figure 4.8: Solutions to and outcomes of problems

identified, sorted in the order of number of users reporting these, highest first. As can be seen, the most common problem was that their software did not support the tasks, or the interface did not make it clear how their task could be performed. The `softwareLimitations` tag was applied to all cases where users reported that they *"would like to do [an activity], but my software does not support it"*, or *"I use software [X], but this particular feature is only available in software [y]"*. The `interfaceDeficiences` tag was applied to all cases where participants reported that *"their device included a particular feature [X], but they never have been able to figure it out."* or *"I do not use feature [X] often because it is very hard to use."* These two conditions were reported in much higher numbers when interacting with mobile devices such as cell phones or PDAs. Since a problem of this type is subjective and arbitrary, it was not used as a representative problem.

The following common problems were simulated in the design of representative tasks in Study 2:

- **Synchronization failed.**
  In the Files task, condition L0 provided no syncing abilities; participants were required to use USB drives or email-to-self to copy files between machines. 12 participants reported this as a problem they had faced.

- **Conflicting versions of information.**
  In the Contacts task, participants were asked to update a person's phone number on a cell phone, and later to look it up on any device they preferred. Several participants failed to

synchronize or to remember that an older version of the data item existed on a different device. 6 participants reported having issues with conflicting versions of information.

- **Duplicate data.**
  In the Files task, participants were instructed that identical copies of their data were placed on two devices, a desktop and a laptop computer. In several cases, when users copied files from one machine to another, they ended up duplicating data. 5 participants reported encountering duplicate copies of data.

- **Forgot to sync.**
  While this is not a problem that was explicitly simulated in the laboratory environment, it occurred as a result of several operations performed by users in the Files and Contacts tasks. 4 participants reported that they had forgotten to sync on more than one occasion.

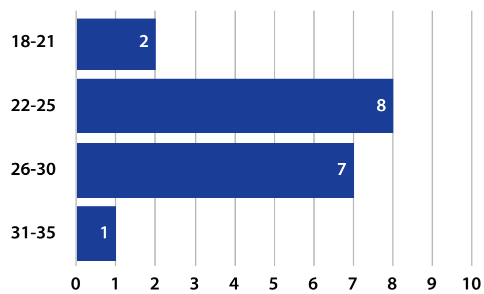
While these are not the top 4 highest-reported problems by numbers, they were representative of broader issues in multi-device computing, and were specific enough to simulate in a controlled laboratory environment.

4.2 Results from Study 2 (Controlled Experiment)

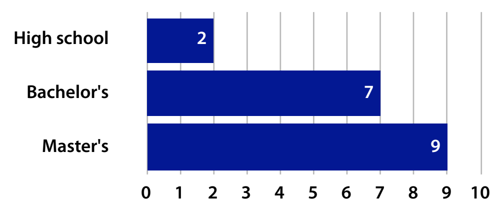
In Study 2, I explored the research questions outlined in section §1.3, using the methods described in §3.4. In this section, I present the results obtained from my experiments, and whether they indicate support or lack thereof for the hypotheses.

4.2.1 Participant Details

A pre-questionnaire was administered to all experimental participants to capture demographic information. 12 participants were male; 6 were female. They were normally distributed in an age range from 18 to 35 (Figure 4.9(a)). 9 had completed a Master's degree, 7 had completed a Bachelor's degree, and 2 had completed a high school degree (or equivalent) (Figure 4.9(b)). All participants identified themselves as either full-time or part-time knowledge workers. Due to limitations with the eye tracking equipment, I was not able to accommodate users who wore spectacles. Among those who participated, 13 participants required no vision correction, while 5 wore contact lenses.



(a) Participants' Age Groups



(b) Participants' Education Levels

Figure 4.9: Participant demographics

4.3 Results for Research Question 1

Research Question 1 explores the impact of (1) different tasks and (2) different levels of system support for migrating information, on the workload imposed on a user.

4.3.1 Overall Workload

From an ANOVA of NASA TLX scores, Task was seen to have a main effect on Overall Workload (OW) ($F_{(2,102)}$ = 4.75; p=0.011). Post hoc analysis using Tukey's HSD showed that the Contacts task imposed significantly lower overall workload than the Files task (p=0.0074). Level of support for performing tasks across multiple devices (L0 vs L1) did not influence Overall Workload and there were no significant interactions.

This suggests that while NASA TLX ratings are able to discriminate between different tasks in the personal information management domain, the scale is not sensitive enough to detect differences in performing a task using two or more techniques. One reason for this could be that NASA TLX, being a subjective measure, can only be administered at the end of a task. It thus fails to capture variation in workload within a task, and provides only an aggregate per-task measure of workload.

Mean scores on the TLX Overall Workload scale were as shown in Table 4.2; ANOVA calculations are in Table 4.3. A comparative illustration is available in Figure 4.10.

Mean (SD)	Files	Calendar	Contacts
L0	41.11 (20.85)	36 (18.80)	30.89 (16.65)
L1	38.61 (18.92)	31.17 (18.91)	22.89 (11.49)

Table 4.2: Means (SDs) of Overall Workload ratings

	DF	Sum of Squares	Mean Square	F-value	p-value
Treatment	1	705	705	2.21	0.14
Task	2	3030	1515	4.75	**0.011**
Treatment:Task	2	137	69	0.22	0.81
Residuals	102	32520	319		

Table 4.3: Overall Workload ANOVA Calculations

Similar effects were seen for three individual dimensions of the NASA TLX scale as well.

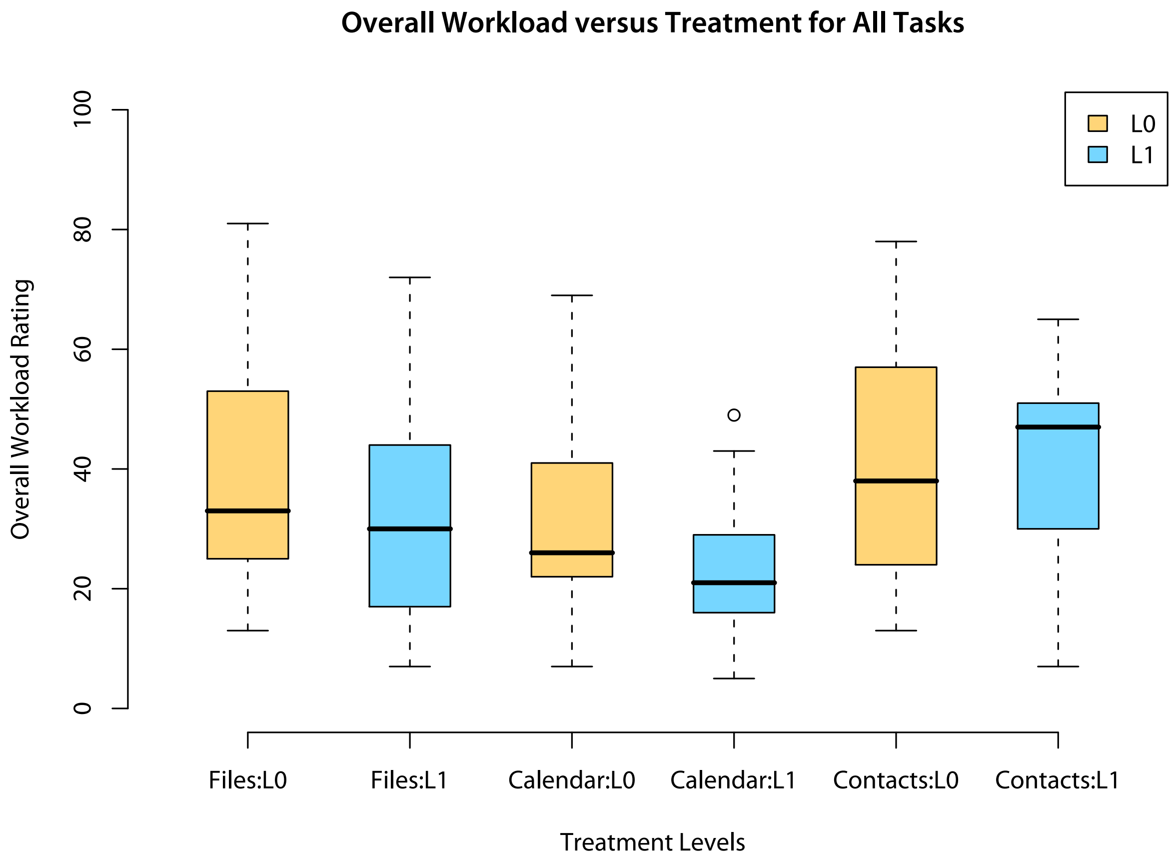


Figure 4.10: Overall Workload across Treatments

4.3.2 Mental Demand

Among individual dimensions of the NASA TLX scale, Task had a main effect on Mental Demand (MD) ($F_{(2,102)} = 6.69$; p=0.0019). Post hoc analysis results for Mental Demand using Tukey's HSD revealed that the Files task imposed significantly higher Mental Demand than the Contacts task (p=0.0024), similar to the effect seen in case of Overall Workload. Treatment level means are presented in Table 4.4, ANOVA calculations in Table 4.5 and illustrated in Figure 4.11.

Mean (SD)	Files	Calendar	Contacts
L0	48.33 (23.89)	41.94 (23.65)	34.44 (15.99)
L1	44.72 (23.36)	44.72 (23.49)	24.72 (11.57)

Table 4.4: Means (SDs) of Mental Demand ratings

	DF	Sum of Squares	Mean Square	F-value	p-value
Treatment	1	334	334	0.77	0.38
Task	2	5837	2918	6.69	**0.0019**
Treatment:Task	2	703	352	0.81	0.45
Residuals	102	44472	436		

Table 4.5: Mental Demand ANOVA Calculations

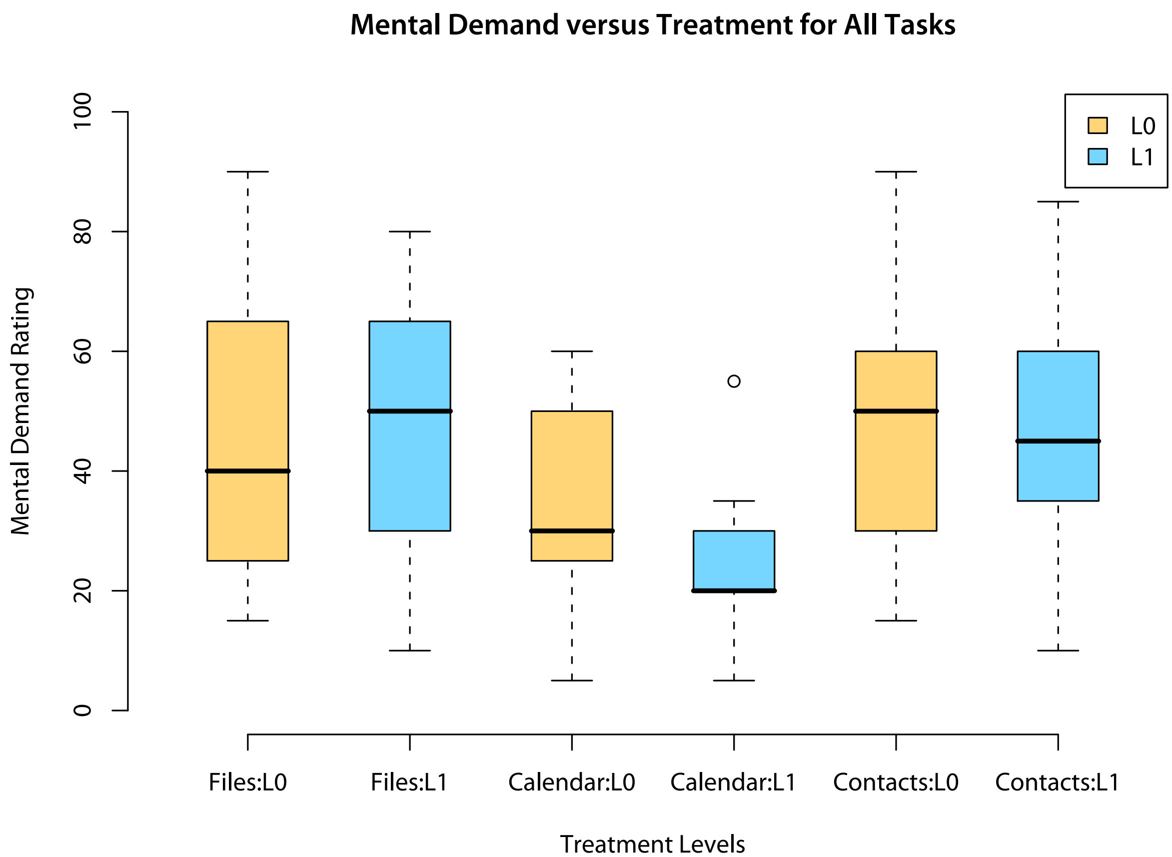


Figure 4.11: Mental Demand across Treatments

4.3.3 Frustration

Task had a main effect on subjective reports of frustration provided by participants ($F_{(2,102)} = 6.57$; p=0.0021). Participants noted significantly higher frustration ratings for the Files task as compared to the Contacts task (p=0.0014, using Tukey's HSD for post hoc analysis). Differences among the other two pairs (Files-Calendar and Calendar-Contacts) were not significant. Treatment level means were as shown in Table 4.6; ANOVA calculations in Table 4.7 the differences are illustrated in Figure 4.12.

Mean (SD)	Files	Calendar	Contacts
L0	43.61 (28.12)	34.72 (25.29)	26.11 (21.32)
L1	39.72 (24.46)	26.39 (21.2)	18.61 (12.34)

Table 4.6: Means (SDs) of Frustration ratings

	DF	Sum of Squares	Mean Square	F-value	p-value
Treatment	1	1167	1167	2.27	0.14
Task	2	6760	3380	6.57	**0.0021**
Treatment:Task	2	100	50	0.098	0.91
Residuals	102	52446	514		

Table 4.7: Frustration ANOVA Calculations

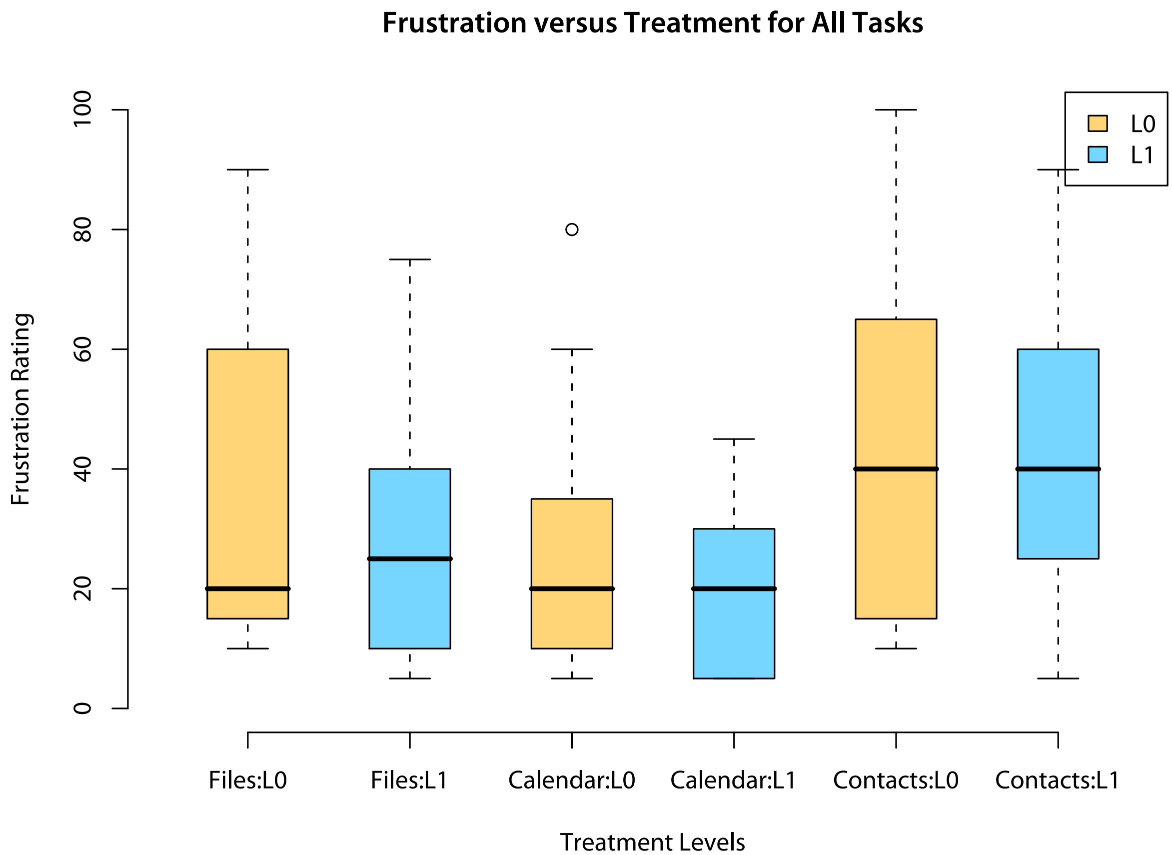


Figure 4.12: Frustration across Treatments

4.3.4 Own (Perceived) Performance

In this dimension, lower numbers indicate better performance. Participants rated their Own Performance differently for the three task conditions, compared using an ANOVA ($F_{(2,102)}$ = 3.37; p=0.038).

Mean (SD)	Files	Calendar	Contacts
L0	30 (24.97)	23.89 (18.67)	18.61 (19.09)
L1	27.5 (21.37)	18.06 (16.55)	15 (17.06)

Table 4.8: Means (SDs) of Own (Perceived) Performance ratings

	DF	Sum of Squares	Mean Square	F-value	p-value
Treatment	1	428	428	1.09	0.30
Task	2	2646	1323	3.37	**0.038**
Treatment:Task	2	52	26	0.066	0.94
Residuals	102	40087	393		

Table 4.9: Own (Perceived) Performance ANOVA Calculations

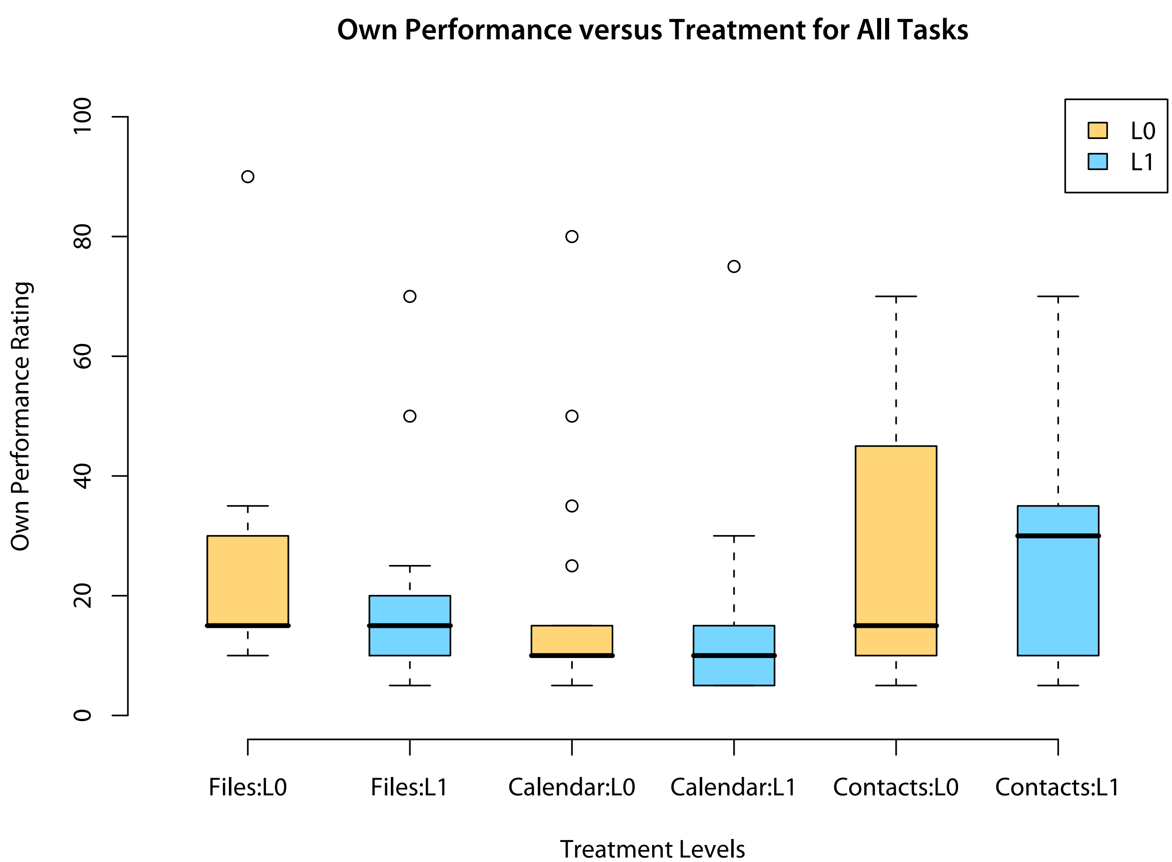


Figure 4.13: Own (Perceived) Performance ratings across Treatments

4.3.5 Other NASA TLX Dimensions

Neither task nor level of support for migration showed any significant differences on the other four NASA TLX dimensions, Physical Demand (Tables 4.10 & 4.11, Graph 4.14), Temporal Demand (Tables 4.12 & 4.13, Graph 4.15), and Effort (Tables 4.14 & 4.15, Graph 4.16).

Mean (SD)	Files	Calendar	Contacts
L0	28.06 (20.30)	28.06 (24.14)	32.5 (23.34)
L1	28.61 (21.34)	19.17 (12.98)	25 (18.47)

Table 4.10: Means (SDs) of Physical Demand ratings

	DF	Sum of Squares	Mean Square	F-value	p-value
Treatment	1	752	752	1.80	0.18
Task	2	587	293	0.70	0.50
Treatment:Task	2	468	234	0.56	0.57
Residuals	102	42579	417		

Table 4.11: Physical Demand ANOVA Calculations

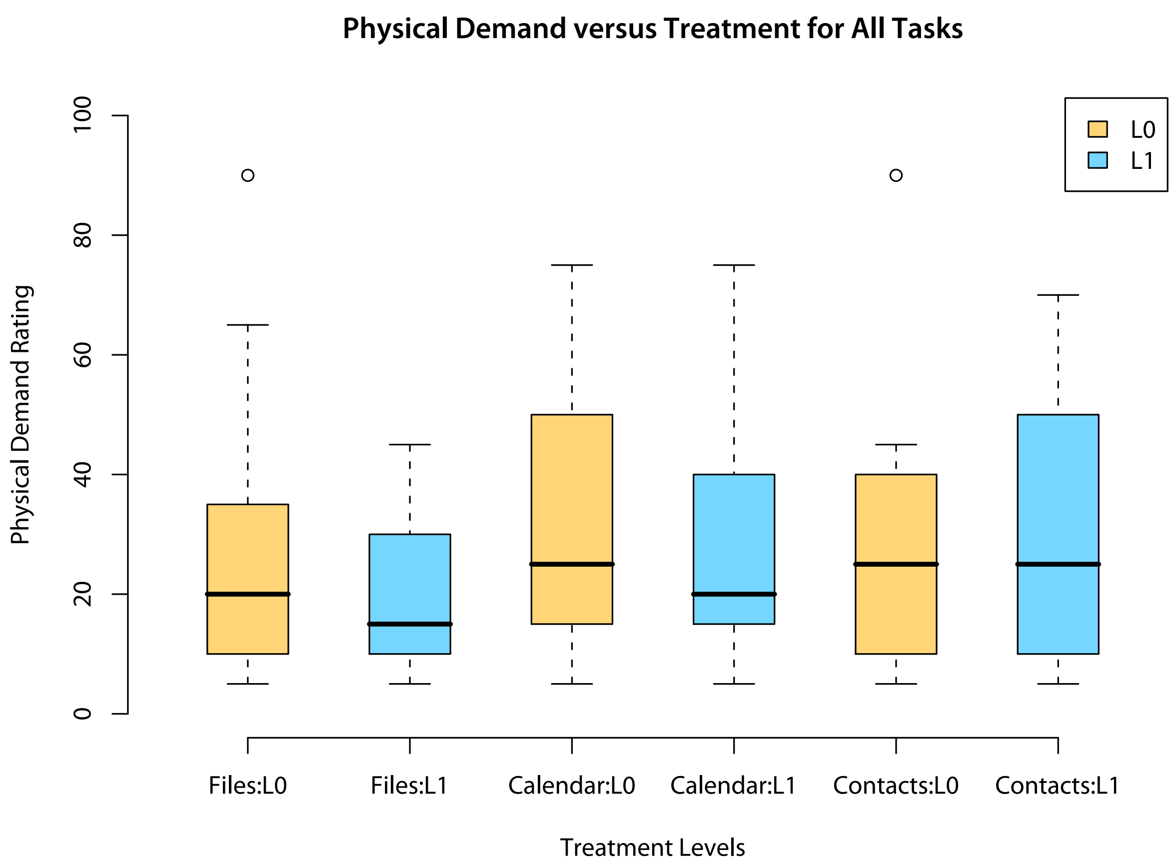


Figure 4.14: Physical Demand ratings across Treatments

Mean (SD)	Files	Calendar	Contacts
L0	41.11 (21.73)	32.5 (21.16)	34.72 (22.26)
L1	33.33 (22.82)	30.56 (22.62)	28.61 (18.85)

Table 4.12: Means (SDs) of Temporal Demand ratings

	DF	Sum of Squares	Mean Square	F-value	p-value
Treatment	1	752	752	1.61	0.21
Task	2	760	380	0.81	0.45
Treatment:Task	2	162	81	0.17	0.84
Residuals	102	47649	467		

Table 4.13: Temporal Demand ANOVA Calculations

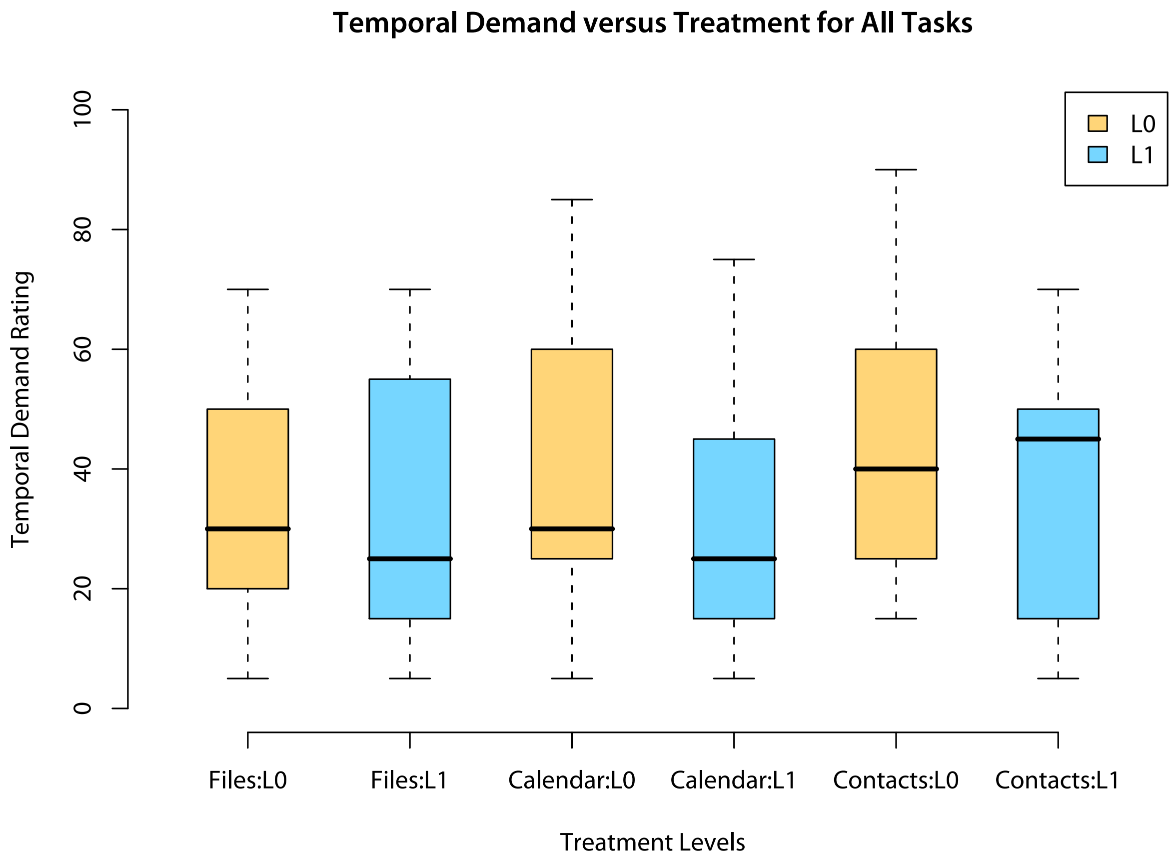


Figure 4.15: Temporal Demand ratings across Treatments

Mean (SD)	Files	Calendar	Contacts
L0	44.44 (28.85)	39.44 (26.95)	36.94 (21.15)
L1	40.56 (24.49)	34.72 (21.86)	24.72 (12.66)

Table 4.14: Means (SDs) of Effort ratings

	DF	Sum of Squares	Mean Square	F-value	p-value
Treatment	1	1302	1302	2.41	0.12
Task	2	2454	1227	2.27	0.11
Treatment:Task	2	379	190	0.35	0.71
Residuals	102	55137	541		

Table 4.15: Effort ANOVA Calculations

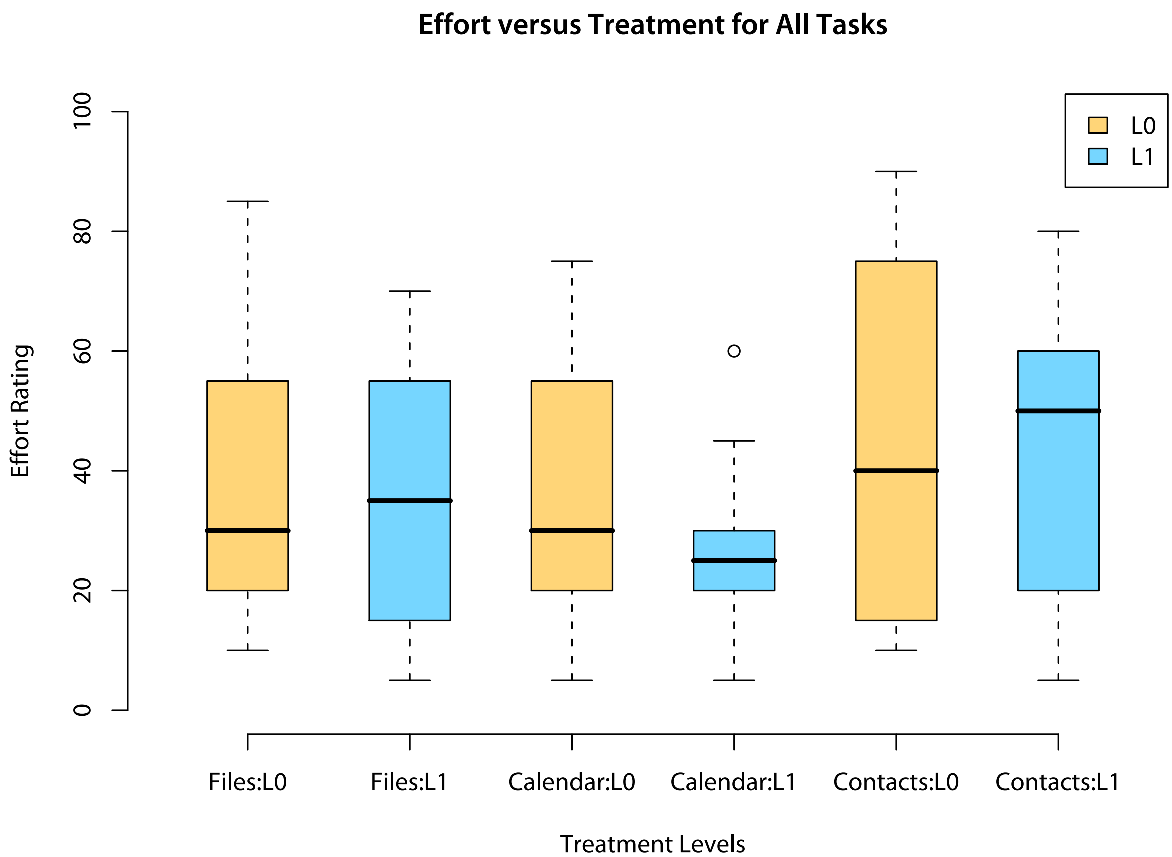


Figure 4.16: Effort ratings across Treatments

4.3.6 Task-Evoked Pupillary Response

In addition to NASA TLX ratings, I also analyzed continuous pupillometric data to examine the effects of task and/or treatment on workload (more details about the method are in §3.9). Since the eye tracker is unable to estimate pupil size when the subject looks at an angle away from the center, or the tracker is improperly fit, no measurements of pupil radius were available for certain intervals for one participant. Accordingly, pupil data for that participant was discarded, though

NASA TLX scores and task performance metrics for this participant were included in the rest of the analysis.

For the Contacts task, significant differences were found for each step between the two levels of system support in task migration (synced versus unsynced conditions.) Table 4.16 shows the means (SDs) and p-values for each step; graph 4.17 illustrates these differences visually.

Step	Mean (SD) for L0	Mean (SD) for L1	p-value
1	12.0 (15.21)	2.89 (7.92)	**0.035**
2	12.16 (13.60)	1.21 (8.16)	**0.0071**
3	13.86 (13.84)	2.15 (8.95)	**0.0062**
4	8.89 (13.63)	0.42 (9.17)	**0.041**
5	13.82 (14.70)	2.94 (8.03)	**0.012**
6	12.67 (13.54)	3.69 (7.41)	**0.02**

Table 4.16: Means (SDs) of adjusted pupil radius for all steps of the Contacts task.

No significant differences were seen for any steps of the Files task (Graph 4.18) or the Calendar task (Graph 4.19).

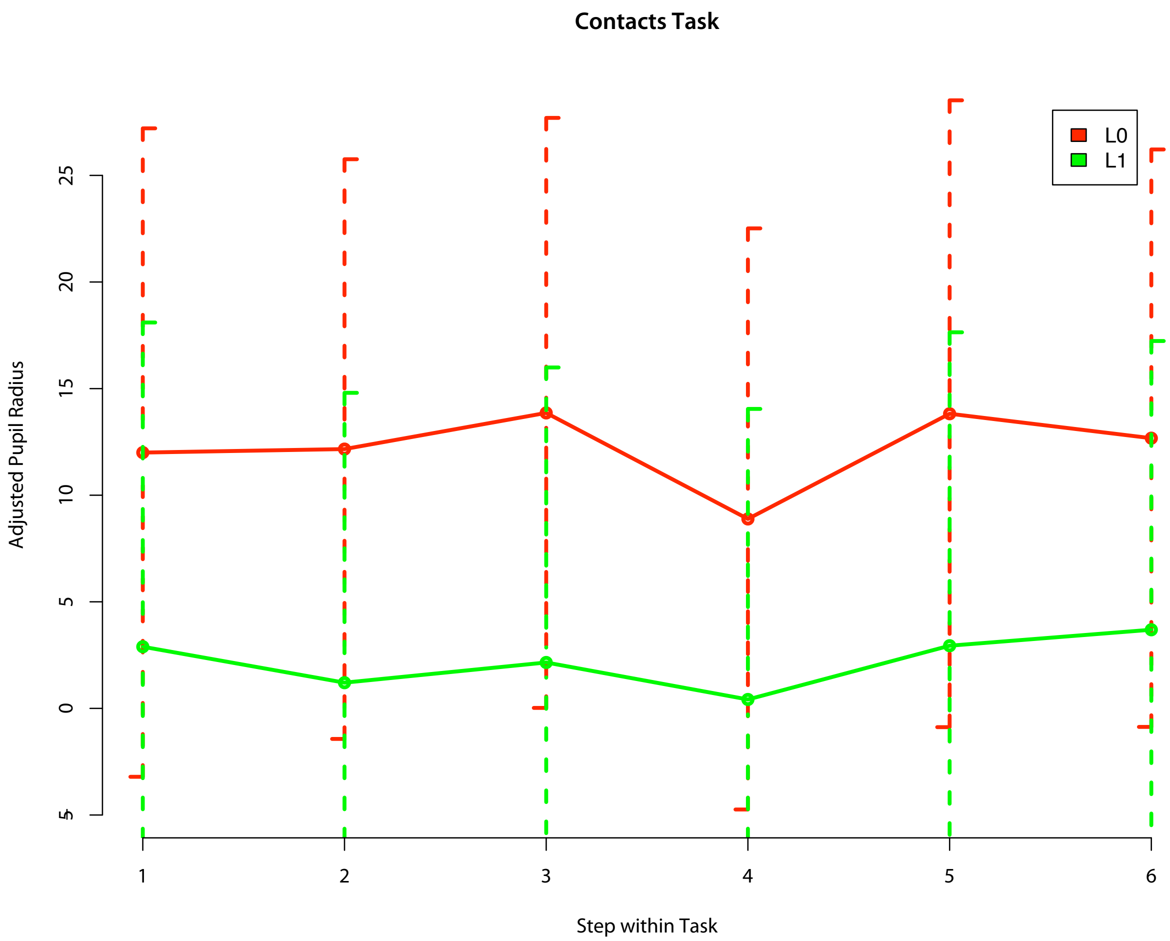


Figure 4.17: Adjusted pupil radius for each step of the Contacts task.

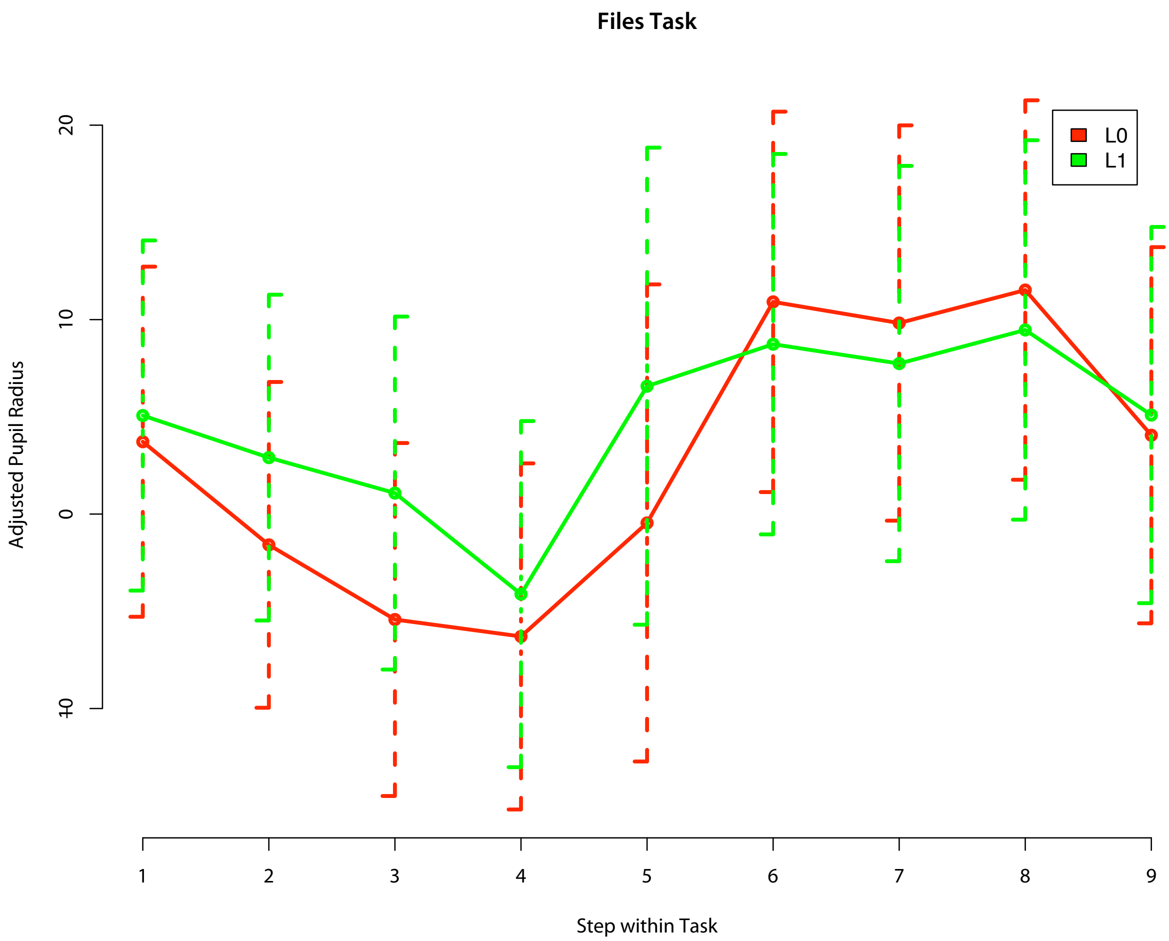


Figure 4.18: Adjusted pupil radius for each step of the Files task.

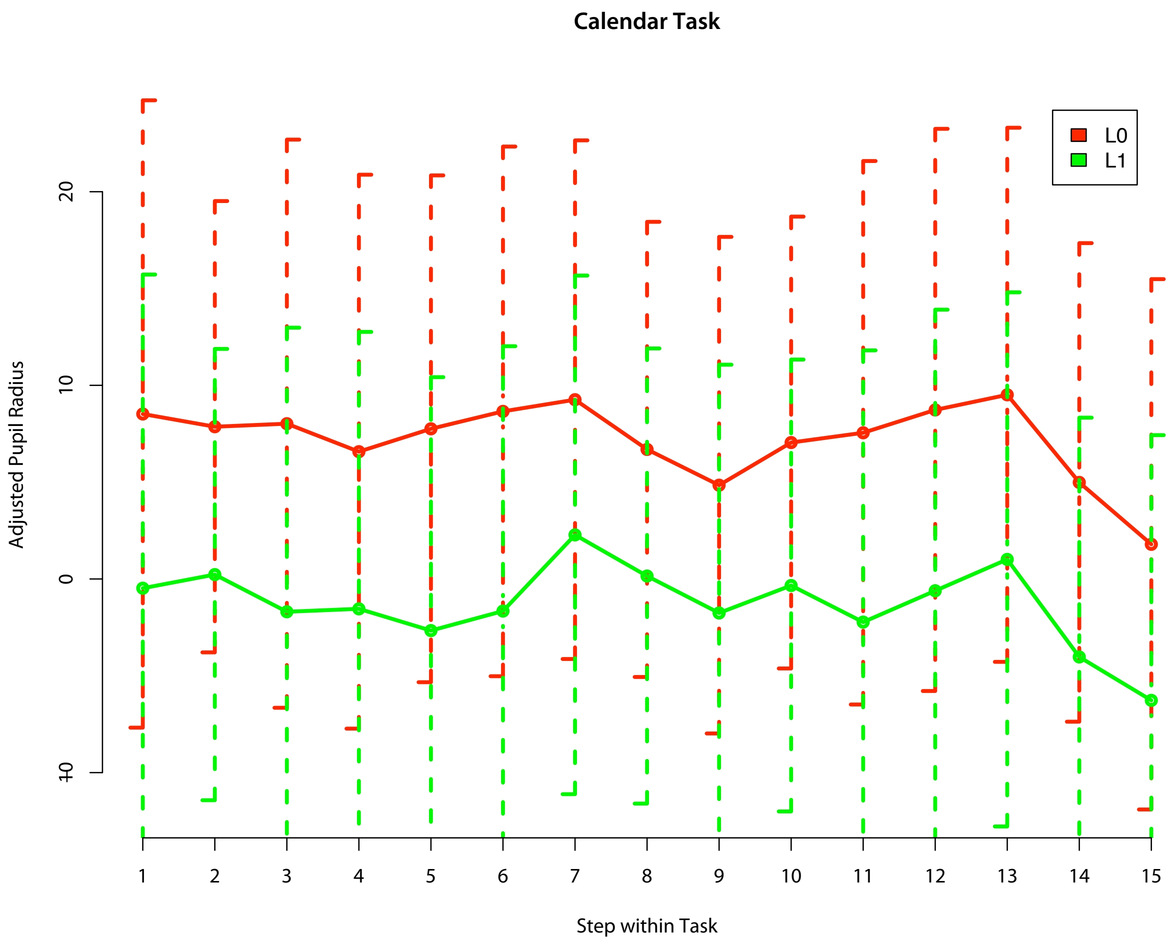


Figure 4.19: Adjusted pupil radius for each step of the Calendar task.

4.3.7 Differences in TEPR Between Steps in the Same Task

In the Files task, Level 0 (where participants used USB drives or email-to-self), significant differences were noted in the workload for the steps before and after the migration step ($F_{(8,136)}$ = 7.8835; p=1.12×10^{-8} using Tukey's HSD). Table 4.17 lists the p-values for all the steps between which significant differences in workload were found.

This suggests that there is a distinct increase in workload before and after the migration step, when there is a lack of support for task migration. It is interesting to note that no significant differences were found in the L1 condition for the same task, suggesting that the file migration support has some effect on differences in workload before/after migration.

	Step 2	Step 3	Step 4	Step 5
Step 6	0.012	0.00018	0.000062	0.028
Step 7	0.032	0.00065	0.00023	-
Step 8	0.0065	0.000085	0.000028	0.016

Table 4.17: p-values for significant differences (Tukey's HSD) for steps before and after migration.

4.3.8 TEPR within Critical Sub-Tasks

Graphs 4.22–4.21 depict the task-evoked pupillary response for several participants for the Files task. These are time-series graphs (time in seconds on the X axis) against adjusted percent pupil radius on the Y axis. In the Files task, Step 5 was the critical task migration step, in which participants were required to pause their task on the desktop and to move to the laptop.

As can be seen, the task-evoked pupillary response (TEPR) rises soon after the start of the critical step, and reaches a (local) maxima. In some instances, it progressively lowers, and in some, it stays at the new, higher level of workload until the end of the task. This provides support for the hypothesis that steps that involve transitions between devices lead to high mental workload.

4.3.9 Summary of RQ 1 Results

In NASA TLX scores, Task was seen to exhibit a main effect on Overall Workload, Mental Demand, Frustration and Own Performance, but not on the other three scales. There was no difference seen on any scale between two treatments levels of the same task. This suggests that NASA TLX is not very sensitive to changes in workload in the kinds of personal information management tasks tested in this experiment. Because of its lack of ability to discriminate between two or more ways of performing the same task, its validity and usefulness in PIM tasks cannot be established with the evidence obtained.

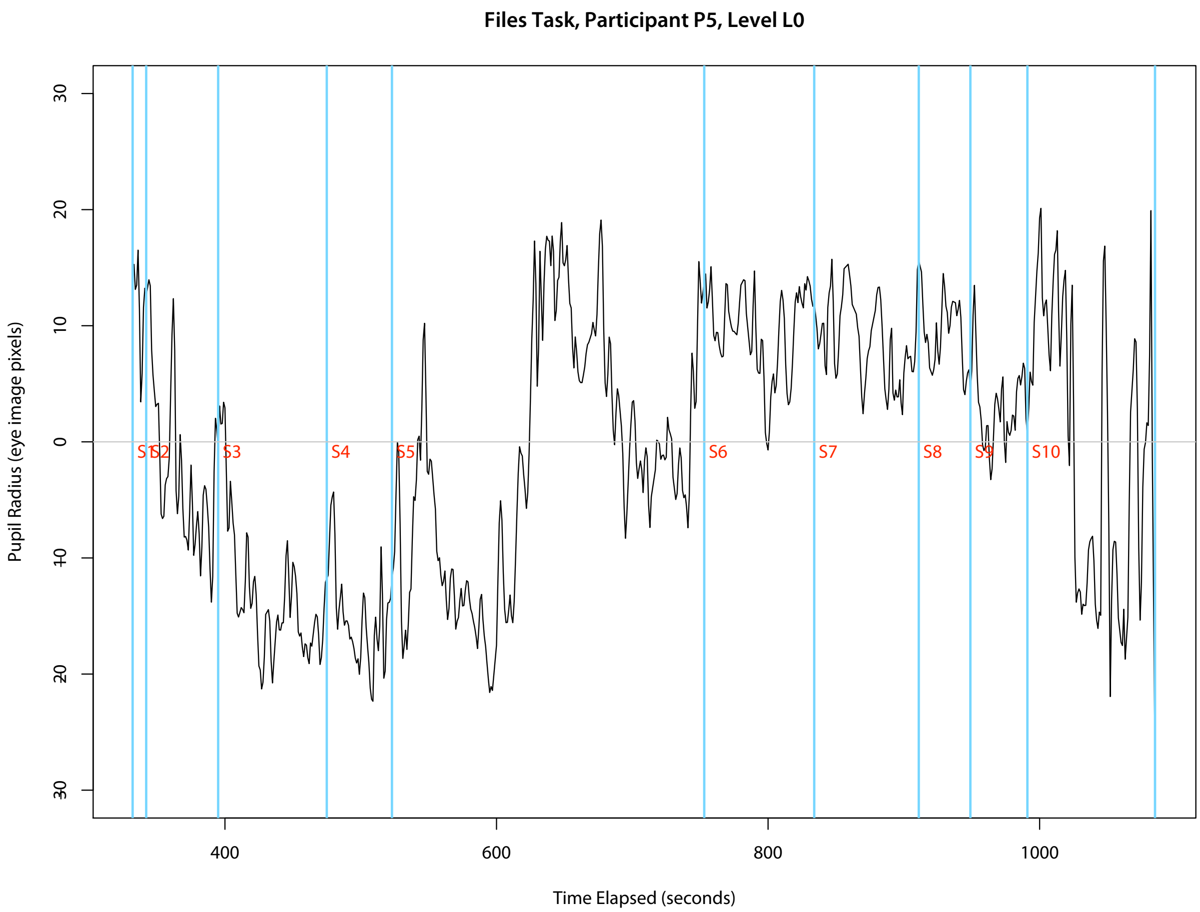


Figure 4.20: Task-evoked pupillary response, Participant P5, Files Task, L0

Task-evoked pupillary response, on the other hand, provided important insights into task migration. Specifically, it showed a significant difference for each step of the Contacts task between levels L0 and L1. Also, it showed significant differences between pre- and post-task-migration steps in the Files task. It was observed from the data that local maximas were attained during the task migration step. All of this points to the potential usefulness of task-evoked pupillary response as a continuous measure of workload in PIM tasks.

Thus, Hypothesis 1 was found to be partially supported, only for specific tasks and specific sub-tasks measured using the task-evoked pupillary response, but not using the NASA TLX scale. Specifically, very few differences were recorded in subjective assessments of mental workload between the two levels of support for each task, but significant differences were noted between different tasks. This suggests that while NASA TLX can discriminate between different tasks, it is not sensitive enough to changes within the execution of each task in this domain. The physiological

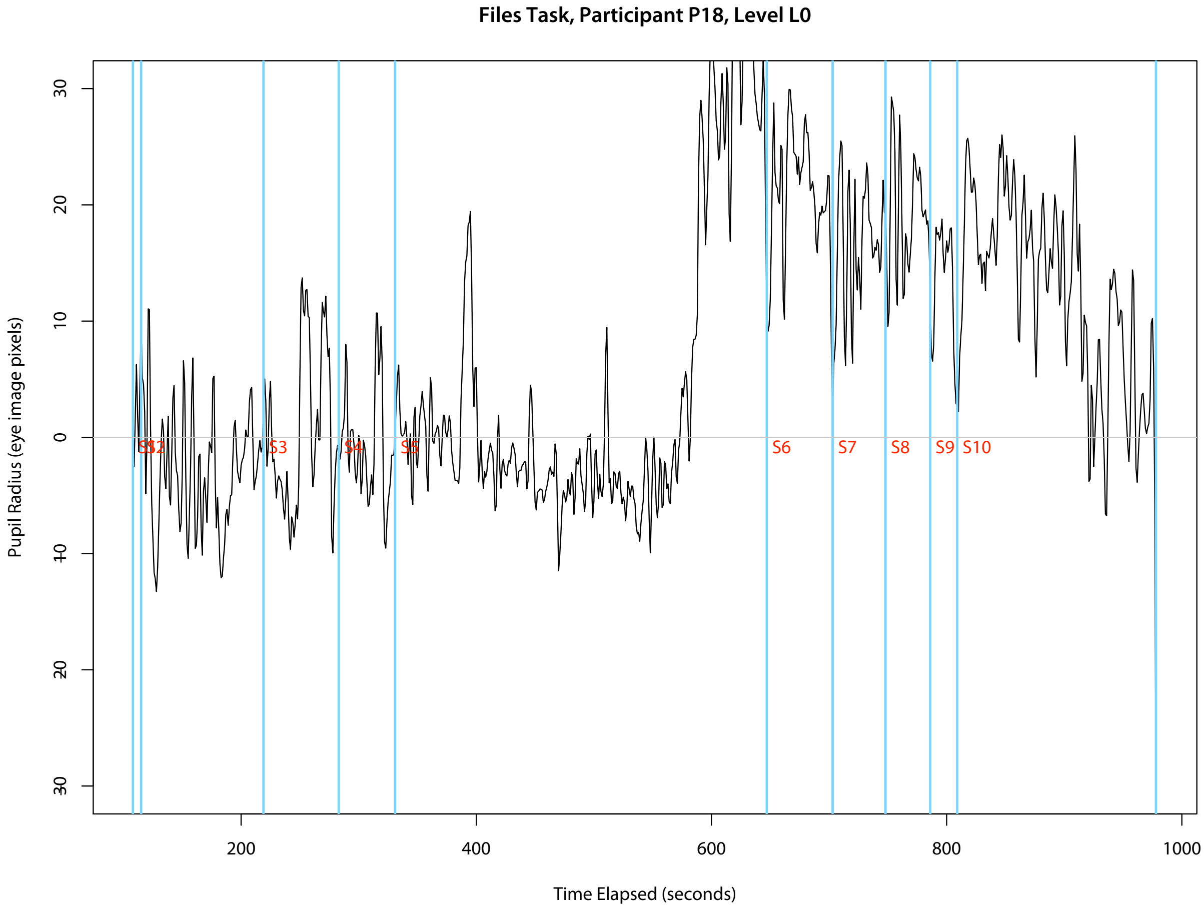


Figure 4.21: Task-evoked pupillary response, Participant P18, Files Task, L0

metric, on the other hand, showed differences before and after the migration step for the Files task, as well as in all steps of the Contacts task. Since this metric highlights intra-task changes in workload that are not detected by subjective metrics, it appears to be a better choice for future workload studies in PIM tasks.

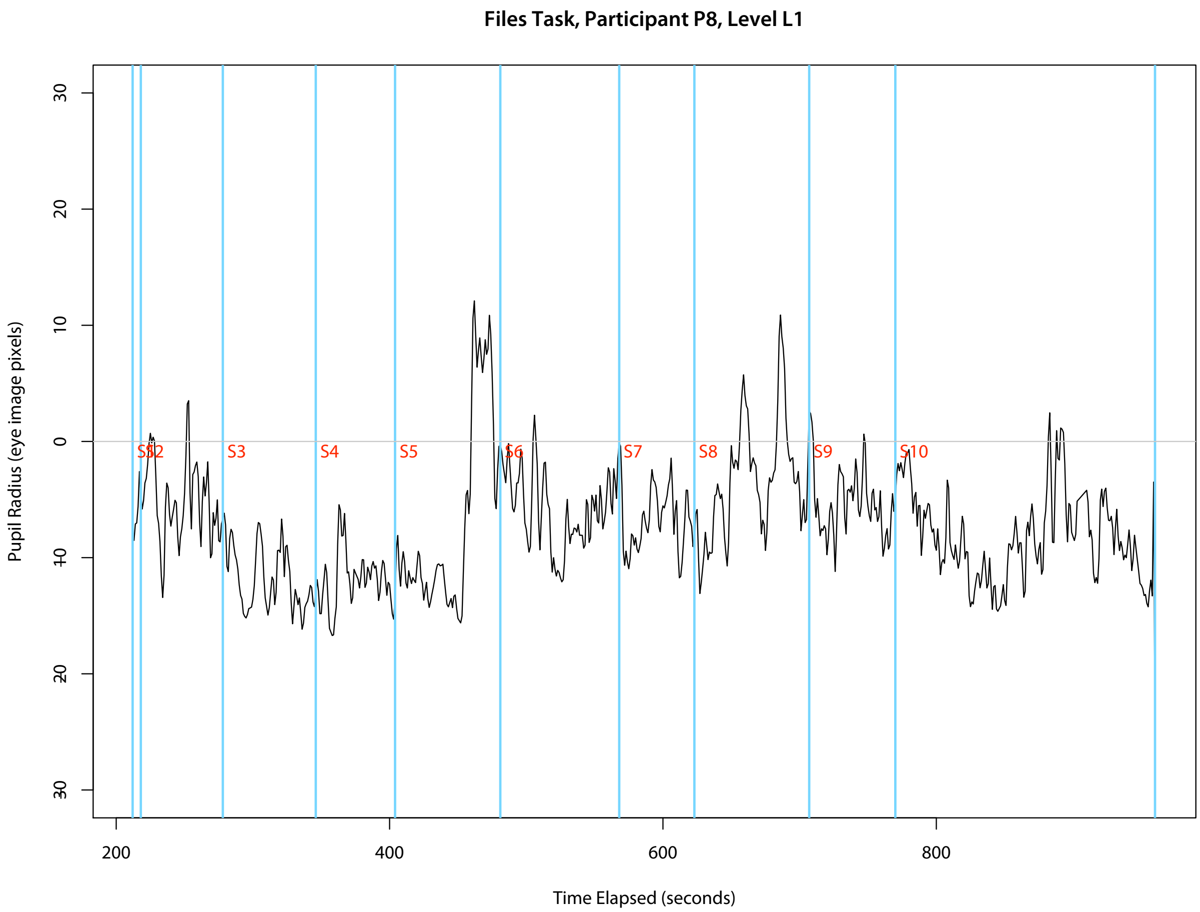


Figure 4.22: Task-evoked pupillary response, Participant P8, Files Task, L1

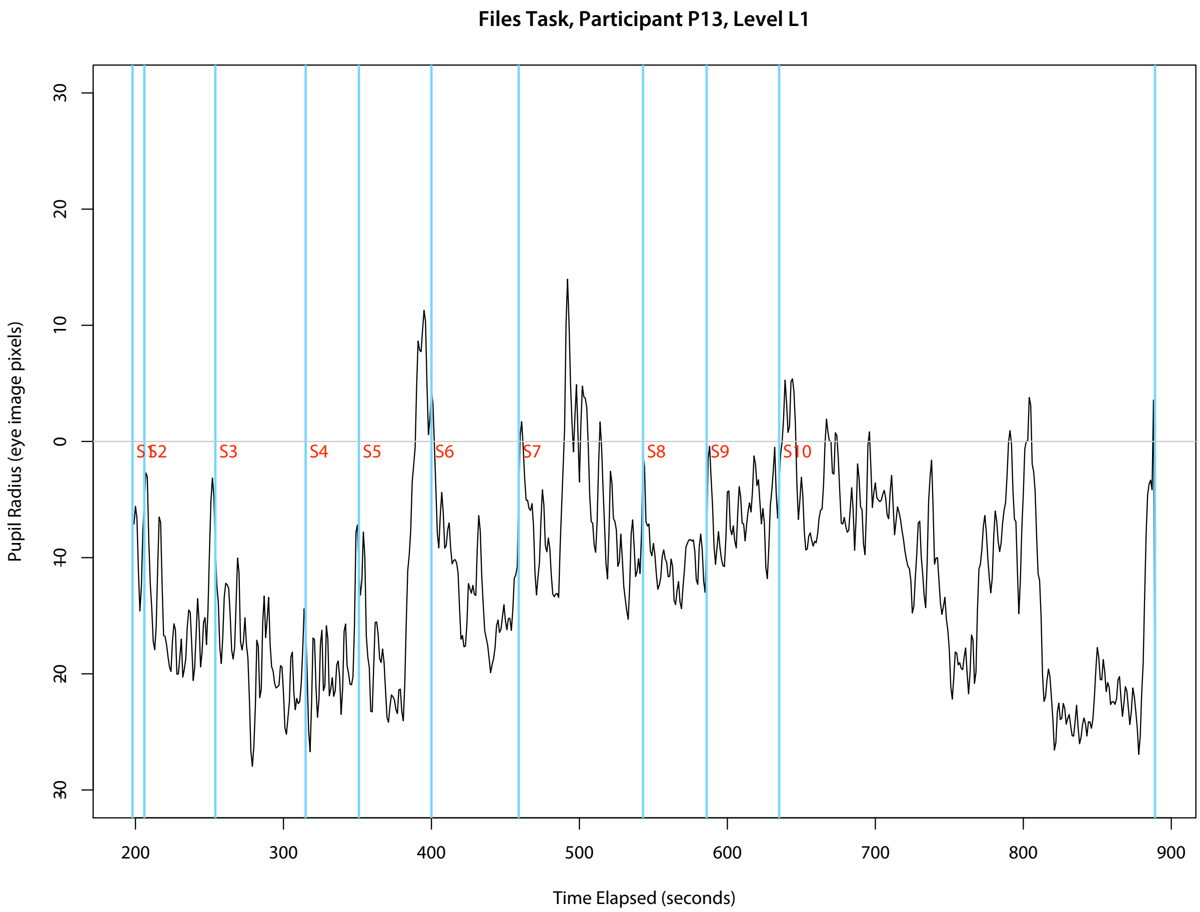


Figure 4.23: Task-evoked pupillary response, Participant P13, Files Task, L1

4.4 Results for Research Question 2

Research Question 2 seeks to explore the differences in operator performance, if any, between the L0 and L1 task conditions. The primary measure of operator performance used in this study (for all tasks) was time on task. Others, such as number of errors, number of entries made, etc. were defined, measured and evaluated on a per-task basis.

4.4.1 Time on Task

Time on task was measured for the entire duration of the task (Overall Task Time), as well as for each step of each task.

Overall Task Time

For the Files and Calendar tasks, no significant differences were found in the time taken to complete the task. However, for the Contacts task, participants completed the task significantly faster in the presence of synchronization support than without ($F_{(1,34)} = 4.72$; p=0.037).

Task	Mean (SD) for L0	Mean (SD) for L1	$F_{(1,34)}$	p-value
Files	2122.33 (873.19)	1850.11 (653.93)	1.12	0.30
Calendar	1905.22 (1196.96)	1992.67 (1070.51)	0.05	0.82
Contacts	2322.78 (1018.65)	1532 (1161.71)	4.72	**0.037**

Table 4.18: Means (SDs) of total time on task for all tasks (in seconds)

Time per Step for each Task

Time on task was then measured and compared for each step in each task, for both levels of system support in migration (shown in graphs 4.24, 4.25, 4.26).

Significant differences ($F_{(1,34)}$=8.83; p=0.0054) were found for the transitioning step in the Files task (Step 5) where participants were requested to pause work on their desktop computers and resume it on a laptop computer, taking their files with them. The mean time taken (SD) in L0 (using USB drives or email to perform the migration) was 312s (276s), while for L1 (using network drives to make the same migration), it was 109s (83s). No significant differences were found for any other step. This was expected; in fact, the lack of significant differences for steps that did not involve a transition from one device to another in the Files task confirms that the experimental setup did not lead to any biases in steps that were identical by design in both treatment levels.

For the Calendar task, two steps took significantly different times in case of the paper calendars versus online calendar ($F_{(1,34)}$=4.33; p=0.045). Both steps involved proposing a meeting time and

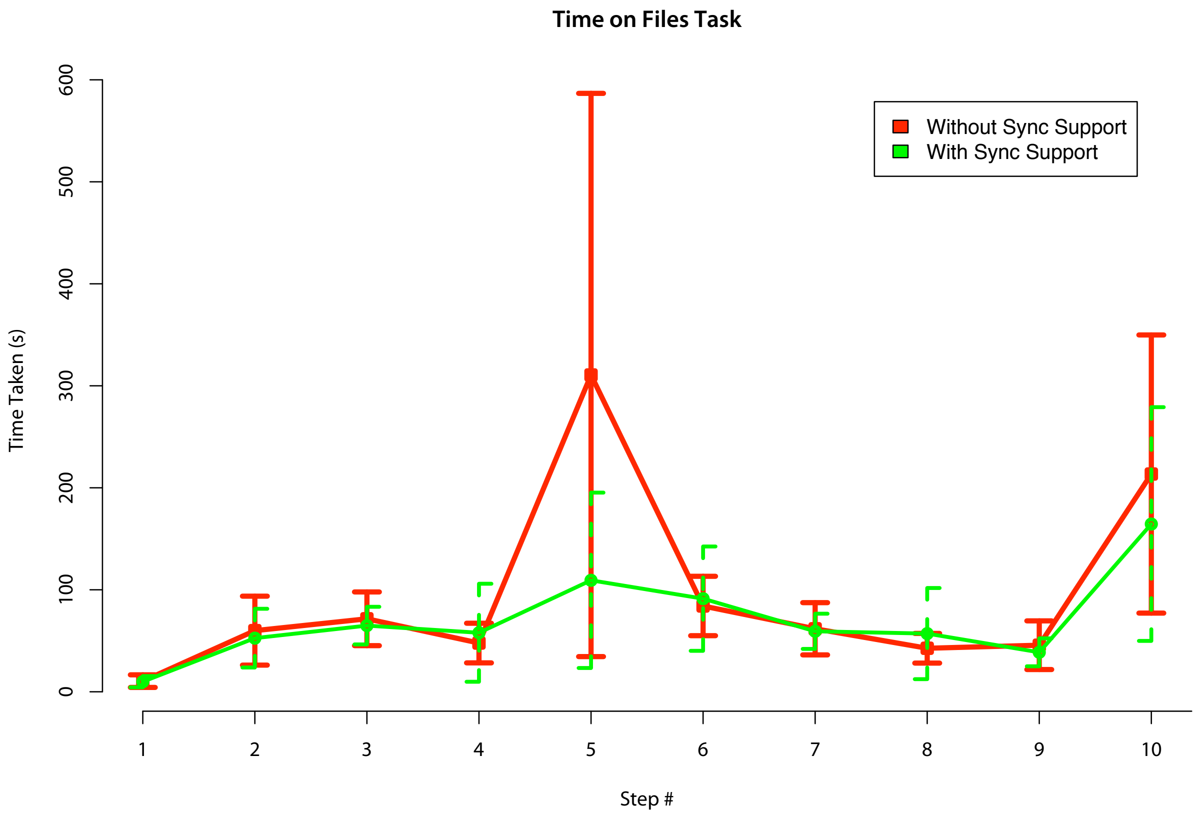


Figure 4.24: Time on task, per Step, in the Files task.

scheduling it on the calendar. In both instances, participants took lesser time using a paper calendar than an online calendar. The ease of quick capture in paper calendars might explain why it is the tool of choice for several users despite the widespread availability of online calendars.

Step	Mean (SD) for L0	Mean (SD) for L1	$F_{(1,34)}$	p-value
2	45.33 (14.48)	55.78 (15.60)	4.34	**0.045**
6	39.28 (12.83)	51.06 (20.79)	4.18	**0.049**

Table 4.19: Means (SDs) of time taken for 2 steps with significant differences in the Calendar task.

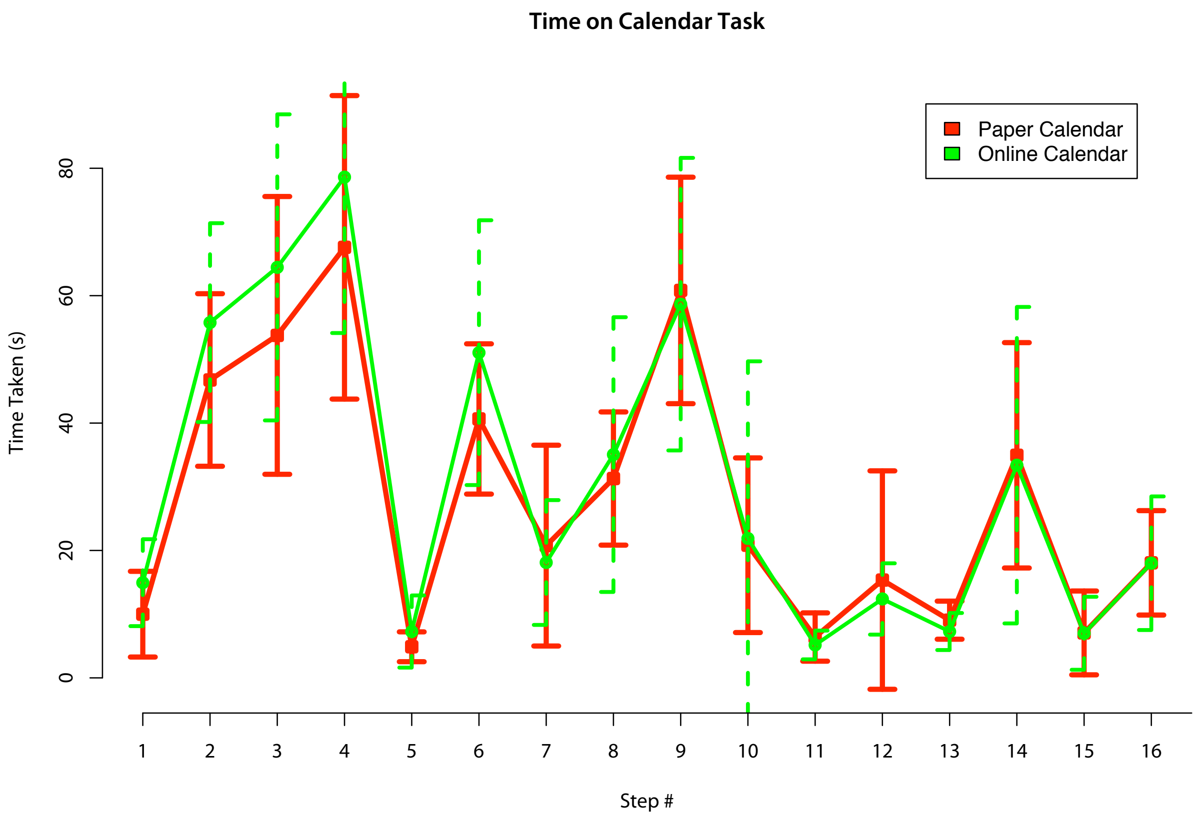


Figure 4.25: Time on task, per Step, in the Calendar task.

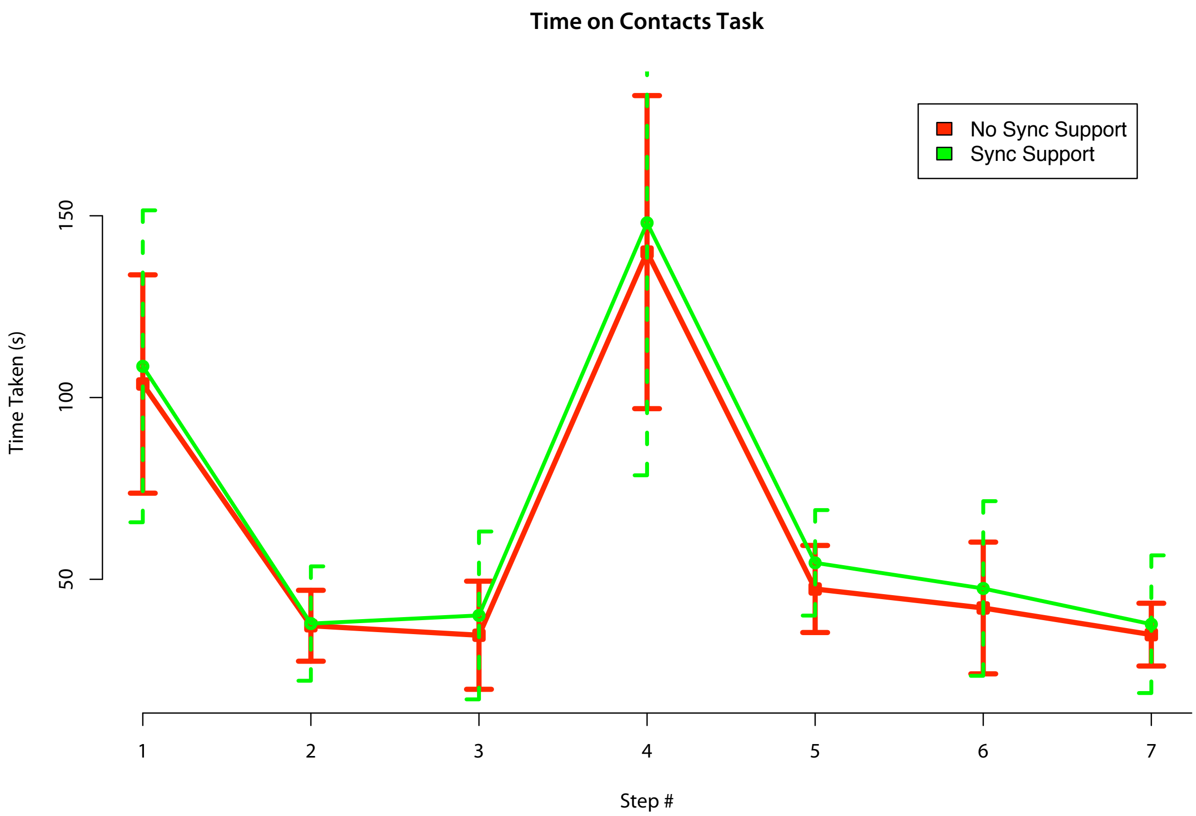


Figure 4.26: Time on task, per Step, in the Contacts task.

4.4.2 Task-specific Performance Metrics: Files

Apart from time on task, several task-specific metrics were taken for each task. For the Files task, the following four metrics were measured:

- Number of files correctly edited;
- Number of files placed in their correct location after migration;
- Number of files copied to the USB drive or to the Network;
- Number of files copied to the Laptop.

While the third and fourth are not task performance metrics per se (i.e., higher numbers do not translate into better performance), I included them in the analysis to examine differences, if any, that might help explain other findings better.

Except the first metric, none were found to have significant effects. It must be noted that since all these metrics were concerned with the status of a limited number of files (14 files manipulated in all), they were subject to ceiling effects. Almost all participants performed these tasks successfully, although some took longer than others — hence the differences in time-on-task, but not on task-specific metrics.

Means (SDs)	L0	L1	$F_{(1,34)}$	p-value
Files Correctly Edited	6.39 (0.92)	5.22 (1.90)	5.52	**0.025**
Files Placed in Correct Location	13.72 (0.83)	13.28 (2.19)	0.65	0.43
Files Copied to USB or Network	10.17 (3.20)	8.78 (5.77)	0.80	0.38
Files Copied to Laptop	13.44 (3.19)	10.61 (5.89)	3.22	0.08

Table 4.20: Means (SDs) for File task metrics

Participants correctly edited more files ($F_{(1,34)}$=5.52; p=0.025) in the condition with no support for file synchronization (Mean = 6.40; SD = 0.92 files) than in the condition with synchronization (Mean = 5.22; SD = 1.90 files) from a maximum of 7 files. This was an unexpected finding, disproving Hypothesis 2 (at least for one particular task metric) that task performance would be higher in the L1 condition.

4.4.3 Task-specific Performance Metrics: Calendar

Two metrics were used in the Calendar task, apart from time-on-task.

- Number of correct answers to schedule-related questions asked during the task performance;
- Number of entries made in the calendar, either paper-based or online.

Neither showed significant differences between treatment levels. Both were subject to ceiling effects because of a cap of 12 correct answers for Q1 and 7 total entries for Q2.

Means (SDs)	L0	L1	$F_{(1,34)}$	p-value
Number of Correct Answers	10.5 (1.71)	11.17 (0.71)	2.35	0.14
Number of Entries Made	5.94 (1.21)	6.33 (0.84)	1.25	0.27

Table 4.21: Means (SDs) for Calendar task metrics

4.4.4 Task-specific Performance Metrics: Contacts

Three additional metrics were used for evaluating the Contacts task.

- Number of correct answers to contact-related questions asked during the experiment;
- Number of entries made on the Primary device; and
- Number of entries made on the Secondary device.

The Primary and Secondary devices referred to are not specifically by hardware, but by their role in the instructions provided. If an instruction clearly required participant to add a contact record to a specific device (either the laptop or the phone), that device was termed the primary device. The other device (either the phone or the laptop, respectively) is then the secondary device.

The number of entries made on the secondary device was significantly different in both treatment levels ($F_{(1,32)}$ = 15.86; p=0.00037): participants who managed contact information with syncing support made 4.71 entries on the other device, while participants without such support made only 1.00 entries. (Table 4.22).

Means (SDs)	L0	L1	$F_{(1,34)}$	p-value
Number of Correct Answers	4.71 (0.85)	5.44 (1.48)	3.16	0.085
Number of Entries Made on Primary Device	8.06 (0.75)	8.06 (0.75)	0.05	0.82
Number of Entries Made on Secondary Device	1 (1.37)	4.71 (3.59)	14.16	**0.00064**

Table 4.22: Means (SDs) for Contacts task metrics

4.4.5 Summary of RQ 2 Results

For the Files task, the time taken to perform the critical step in the Files task — moving from the desktop to the laptop — was significantly higher when there was a lack of system support for such migration (implemented in this experiment as a Network Drive). However, more files were edited correctly in the case where synchronization had to be performed using USB drives or email-to-self. For Calendars, there was no difference in any task metrics between the paper and online calendar conditions. In the Contacts task, more entries were recorded on secondary devices when synchronization was automatic.

Thus, little to no support was found for Hypothesis 2, especially with the observation that more files were edited correctly with lower levels of support for task migration.

4.5 Results for Research Question 3

Research Question 3 examines if measures of mental workload may be used as predictors of task performance in personal information management tasks. Since time-on-task was the only performance metric that was (1) used for all three tasks, and (2) was not subject to any ceiling effects, further analysis of the correlation between performance and workload focuses on this metric. Mental workload was estimated via two methods; we consider them separately to examine whether either or both of them may be used as task performance predictors.

4.5.1 NASA TLX Ratings as Predictors of Operator Performance

Little to no correlation was seen between NASA TLX sub-scales with task performance measured as time-on-task. Pearson's product moment coefficients (for Overall Workload × Time on Task) are as in table 4.23.

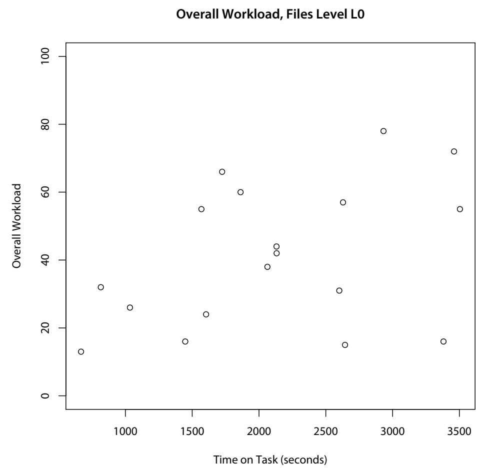
Overall Workload	L0	L1
Files	r=0.39, p=0.11	r=0.57, p=**0.01**
Calendar	r=0.19, p=0.44	r=-0.017, p=0.95
Contacts	r=-0.33, p=0.18	r=0.025, p=0.92

Table 4.23: Pearson's r values for Overall Workload in all task conditions.

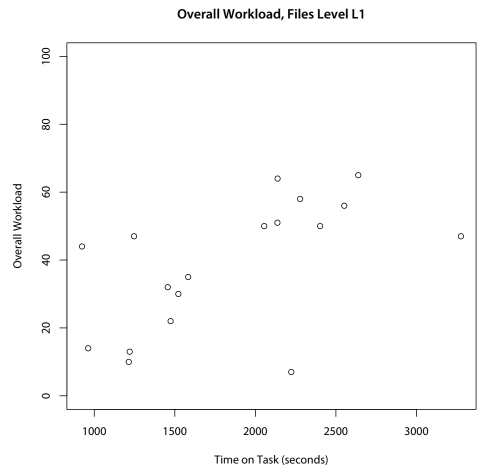
Among all tasks, significant correlations were seen in the following cases:

- **Overall Workload for Files Level L1.**
  $p = 0.01$, $r = 0.57$
- **Mental Demand for Files Level L1.**
  $p = 0.0071$, $r = 0.61$
- **Own (Perceived) Performance for Files L0.**
  $p = 0.05$, $r = 0.47$

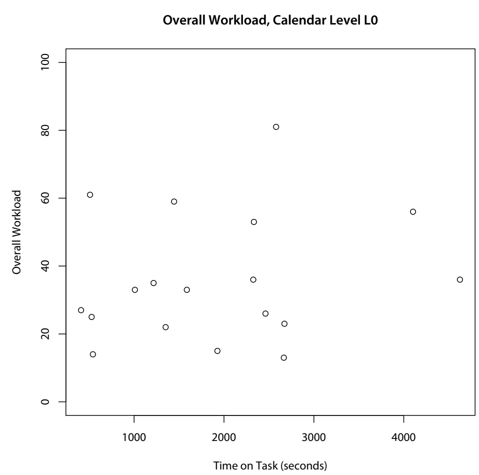
- **Own (Perceived) Performance for Files L1.**
  $p = 0.02$, $r = 0.54$
- **Frustration for Files L0.**
  $p = 0.05$, $r = 0.47$
- **Frustration for Calendar L0.**
  $p = 0.51$, $r = 0.17$



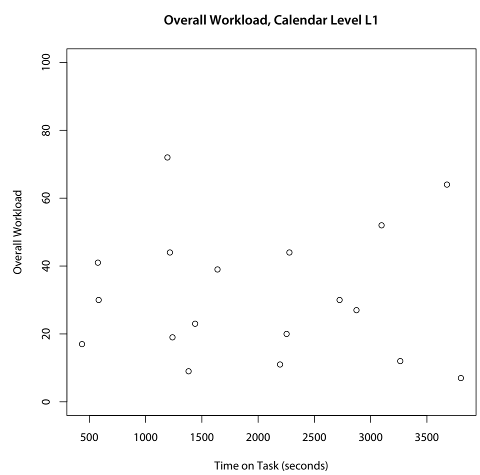
(a) Files L0



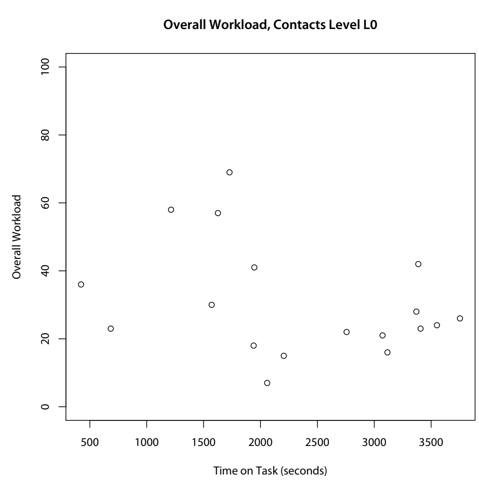
(b) Files L1



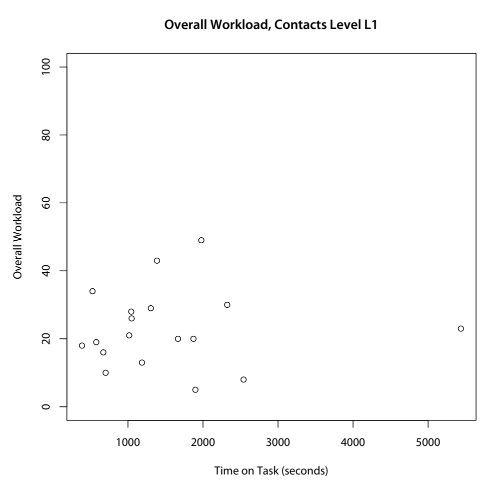
(c) Calendar L0



(d) Calendar L1



(e) Contacts L0



(f) Contacts L1

4.5.2 Task-Evoked Pupillary Response as a Predictor of Operator Performance

Workload estimated according to the Task-Evoked Pupillary Response was not found to be significantly correlated with Time on Task, using Pearson's product-moment coefficient (r). Table 4.24 shows the correlation coefficients and p-values for each task condition. It can be inferred that mental workload (measured via pupillary response) is not a good predictor of task performance.

TEPR × Time for each step	L0	L1
Files	r = -0.062, p = 0.46	r = 0.15, p = 0.063
Calendar	r = -0.11, p = 0.078	r = -0.067, p = 0.283
Contacts	r = -0.13, p = 0.18	r = 0.042, p = 0.68

Table 4.24: Pearson's r for Task-Evoked Pupillary Response for each task condition.

4.5.3 Summary of RQ 3 Results

Neither NASA TLX ratings nor task-evoked pupillary response showed consistent correlation with task performance. Isolated instances of significant correlations were observed, but they do not support the use of workload measures as predictors of task performance. The lack of any meaningful correlation between pure performance-based metrics and workload metrics suggests that neither alone is sufficient to assess and describe highly contextualized tasks in the domain of personal information management. Thus, Hypothesis 3 was disproved in case of both metrics used in the measurement of mental workload.

4.6 Interesting Observations

While the preceding sections provide answers to the research questions posed at the start of this study, there were several interesting observations I noted while participants performed the experimental tasks. These are categorized and summarized here, with implications for the design of systems that may be able to avoid or minimize the impact of some of these issues for users. These observations were not hypotheses, hence are not statistically treated; however, they are included here for completeness.

4.6.1 Preparation (or Lack Thereof) in Task Migration

All participants were informed in the instructions preceding the Files task that they would start their task on the desktop, and would be requested mid-task to pause and resume their activity on a laptop. To provide an opportunity to participants to perform any planning operations before beginning the actual tasks, the first instruction I1 specifically requested them to *'proceed to [their] desk and settle down'* and to *'let [the experimenter] know when [they were] ready to begin.'*

It is interesting to note that none of the participants took this opportunity to plan for the upcoming transition. Since the means of task migration (USB drives, email access and network drive access) were already provided to them, it would have been possible for them to plan ahead by copying their files to the network, for example. However, none did so. Even during the transition task, several users declared that they were done and ready for the next task (on the laptop) when they clearly were not. E.g. if they had copied their files to the USB drive, but had not copied those files from the USB drive to the laptop, clearly they were not ready to begin the tasks yet. However, when presented the instruction for the next task (to edit a specific file), they then proceeded to do the second half of the migration steps.

This lack of planning has significant implications for those designing technologies for mobility: users cannot be expected to plan ahead or to prepare for a device transition [Perry et al., 2001]. Task migration technologies must take into account the opportunistic use of multiple devices without any pre-planning and must initiate any pre-migration activities without the need for explicit user intervention [Pyla et al., 2009]. Users' desire to switch from one device to another cannot be expected to be expressed ahead of time, so systems must instead interpret the intentionality and take appropriate action.

4.6.2 Aversion to Manual Syncing

While performing the Contacts task with synchronization capabilities (treatment level L1), several participants did not partake the offer. Even after being explained how to sync at the beginning of the task, and offered the choice to sync any time, many chose not to sync anyway. This observation contrasts with the Files task, where all participants synchronized their data during the transition. I hypothesize that this difference in behavior is because of the lack of a motivating factor in case of Contacts, and the perception of a 'safety net' in the knowledge that their cell phone as well as laptop will be available at the time of performing lookups. If other means of lookup are expected to be available, (e.g. being able to check the phone as well the laptop when needed), participants did not appear to be motivated to keep their information consistent across devices; they only needed to ensure access to it in some form or another.

This, coupled with the tendency of participants to lookup information on the laptop rather the cell phone, led to a few incorrect answers. Specifically, participants looked up incorrect outdated information on the laptop without realizing that an updated copy was present on the cell phone. The system provided no indication to the user about the likely staleness of the information on the laptop. In other cases, when a lookup failed on the laptop (e.g. entry not present), participants performed the lookup a second time on a different device (i.e., the cell phone) and were able to answer correctly.

For designers of multi-device systems, the implications are obvious: users are not likely to perform task migration activities unless there is an immediate perceived benefit in doing so, or the risk of failure. They are likely to access stale data unless there is a method to provide contextual notifications about the status of this data.

4.6.3 Maintaining Contextual Awareness in Calendars

In the Calendar task, a few of the instructions provided to the participants mentioned the current date as a way to anchor them in temporal context. Since an entire week's worth of calendar events were presented in about 10 to 15 minutes, it was important to introduce the current day in order to preserve the hypothetical temporal unfolding of events in the experimental tasks.

Participants adopted various techniques to maintain this temporal context while interacting with the calendars. Those who used the electronic calendar clicked the specified date in the calendar window, which would then highlight that day in the display. Such a visual representation helped as an external cognition aid so that the task of remembering the current day could be offloaded to the environment. Very few users who used paper calendars used similar techniques: those that did, marked each passing day with a dot or a cross towards the top of the day.

I hypothesize that the permanence of any markings made on paper might have been a contributing factor towards the decision not to make any temporary contextual marks on it. On the other hand, the ephemeral nature of the highlighting in the electronic calendar provided a degree of awareness without imposing a permanent record of the activity.

4.6.4 Capturing Information about Tentative Events in Calendars

The scheduling of tentative collaborative events caused a high amount of confusion to users (noted via experimenter's observations; not statistically significant). Using multiple paper calendars, participants indicated the changes and rescheduling with an assortment of arrows, scratched lines, and other idiosyncratic annotation techniques. In electronic calendars, while participants could reschedule an event easily by dragging-and-dropping the electronic representation of the event to the rescheduled time, this did not solve the entire problem.

The larger issue in tentative collaborative events is the ad hoc specification of attendees' constraints. Current calendar systems do not capture the set of constraints that lead to the tentative scheduling of an event. Hence, when such an event is to be moved to another time, the new start time must be evaluated against the complete set of constraints by consulting the originating source, e.g. email. The event record within an electronic calendar provides no way to indicate the justification behind the particular choice of time, and thus lacks an affordance for potential rescheduling. This is also a problem when adding a new constraint to the mix.

While a few calendar systems do provide support for automatic multi-party meeting scheduling, the resulting artifact is a calendar event, not an expression of the constraints. This makes it difficult to add or remove constraints from the mix, to arrive upon a different time than originally scheduled.

The direct implication from this observation to designers of calendar systems is to provide a way to capture these constraints such that events may be rescheduled without having to start from step one, reading the (potentially long) series of emails that prompted the meeting request.

Chapter 5

Discussion

> "We might be measuring things right, but are we measuring the right thing?"
>
> — *Quote adapted from [Drucker, 2006]*
> *With apologies to Peter Drucker.*

Through the results of these studies, I found that specifics of the tasks and levels of support for task migration affected users' perceived workload ratings as well as task-evoked pupillary response in a variety of ways. Certain tasks were rated as frustrating to a higher degree than others, or elicited higher mental demand. In specific sub-tasks, I also saw an increase in task-evoked pupillary response. An effect was seen in the time they required to perform certain tasks, though only for the critical steps in each task, and not consistently across tasks. These workload metrics were not the traditional usability metrics that are often used to evaluate computing systems, such as performance, efficiency, errors, etc. In fact, traditional metrics such as whether users were able to answer questions correctly and time-on-task showed little to no difference with the different ways of performing a task, with and without support for task migration.

What this points to is that while both types of systems result in similar outcomes (and thus would be rated equally on traditional usability metrics), they do not evoke the same experiences in users. Frustration, mental demand, and workload: all are components of the entire user experience, but are not often captured by researchers and designers when assessing personal information ecosystems. This points to two separate, yet related, issues that warrant discussion: (1) evaluating usability using concepts from hot cognition that are more representative of user concerns when using multiple devices together, and (2) evaluating usability for a device ecosystem together instead of as disparate devices.

5.1 Evaluating Usability using Hot Cognition Aspects

Besides the need to measure traditional usability metrics, it is important to test whether we are, in fact, measuring the right metrics — whether they matter to the user experience, or simply are indicative of first-paradigm thinking [Harrison et al., 2007] where third-paradigm thinking is more appropriate. Dillon notes [Dillon, 2002a] that in several tasks, efficiency may not be the user's priority. In particular, he highlights the inadequacy of traditional usability measures for many high-level, ongoing tasks such as information retrieval and data analysis.

Other studies also have shown [Park et al., 2006] that users' preferences for particular brands of devices have significant effects on their perception of usability of those as well as other devices. This shows that aspects of hot cognition such as affect, emotion, personal preferences, etc. play an important role in the user experience — perhaps an even greater role than purely objective metrics such as task completion times and feature comparisons.

5.2 Holistic Usability for Personal Information Ecosystems

Distributed cognition theory recognizes that actors in a system often rely on the use of external artifacts to augment their own cognition. Usability cannot thus be embedded into an artifact, but is distributed across an entire activity system [Spinuzzi, 2001]. This is evident in this study in various ways: users performing the Calendar task kept track of the current day by highlighting that day in an online calendar, or by marking off corresponding days in a paper calendar. In the Files task, a few users kept modified files open in their respective editor programs as a means of tracking their changes. While these are just a few idiosyncratic examples, it points to the larger issue of systems and devices lacking explicitly-designed support for external cognitive tasks.

In Study 1, many survey respondents considered the presence of too many features in a device as a liability than an asset (section §4.1). But most often, these are the metrics that are touted on product specification sheets and in advertisements. When product usability evaluations are conducted for each device separately, they fail to account for the use of the device in a broader context of use, nestled among other devices with which it must interface [Pérez-Quiñones et al., 2008].

This work already has begun, but it must go further on. Mills and Scholtz [Mills and Scholtz, 2001] describe an approach to multi-device design, situated computing. Specifically, they stress the need to *"remove the tyranny of an interface per application, per device."* Rekimoto [Rekimoto, 1997] describes a multi-device interface that allows the user to drag items from one machine and drop them into another. Earlier, I developed the Syncables framework [Tungare et al., 2007] that permits users and applications to seamlessly access their information on any device, irrespective of type, format, or the device on which it currently exists. These examples, although sparse, highlight some of the paths that may be taken to achieve true multi-device interfaces of high usability.

Chapter 6

Conclusions & Future Work

6.1 Conclusions

In this dissertation, I examined the problems that users face when managing personal information using multiple devices such as laptop computers, desktop computers and cell phones. Through my first study, I learnt that several people encountered various kinds of difficulties in using multiple devices. Some of them chose to forgo the luxury of making context-appropriate device choices, and instead opted for a single device to minimize the need for task transitions. Motivated by these initial findings, I conducted a controlled laboratory experiment to study this further.

In the second study, participants performed 3 tasks, related to Files, Calendars and Contacts, using desktops, laptops, and phones, at two levels of system support for multi-device interaction. It was important not only to measure their performance on these tasks, but also to understand their perceptions and mental workload while they performed the tasks. They completed the NASA TLX subjective workload assessment for each task, and I obtained a physiological measure of workload in the form of the Task-Evoked Pupillary Response measured using an eye tracker.

Ratings on 3 of the NASA TLX sub-scales and Overall Workload showed differences between tasks, but none of the scales were able to discriminate between two conditions in the same task. This suggests that NASA TLX does not appear to be a sensitive test in the domain of personal information management. On the other hand, the continuous measure of workload, Task-Evoked Pupillary Response, was able to detect changes at the sub-task level, significant differences before and after migration in a specific task condition (Files task, without support for file synchronization) as well as differences in every step of the Contacts task in the two conditions. All of this suggests that workload estimated from pupil radius shows promise in this area to evaluate tools and systems from a hot cognition point of view.

The time taken to perform the critical step in the Files task — moving from the desktop to the laptop — was significantly higher when there was a lack of system support for such migration (implemented in this experiment as a Network Drive). However, participants edited more files correctly without synchronization support, an unexpected finding. In the Contacts task, participants entered more information on their secondary device (i.e. the device that was not explicitly mentioned in the instructions) when synchronization support was available, suggesting that participants are not likely to copy data between devices without a strong motivating reason. No support was found to indicate that workload ratings correlated with task performance, as has been found in several other domains.

In addition, I noted several interesting observations in the notes I took during each experiment. None of the participants prepared for the upcoming transition in the Files task, even when they were aware of it, and were provided an opportunity to do so (details in §4.6.1). In the Contacts task, when provided the means to synchronize contacts with a single button press, many participants did not avail of this feature (§4.6.2). Participants used a variety of techniques to keep track of the current day when using calendars, but a few were wary of making permanent marks on their paper calendars for such temporal contextual data (§4.6.3). In the Calendar task, for steps that involved tentative scheduling with collaborators, participants encountered difficulties changing or adding constraints to the mix (§4.6.4). Current calendars do not provide adequate support for noting attendees' constraints when scheduling group events, and thus, make it tough to find an alternate suitable time when attendees' availability changes.

6.2 Contributions

Pure performance-based measures are not sufficient to describe and assess highly contextual tasks in the domain of personal information management, and the inclusion of user perception in their assessment is important. Traditional usability metrics emphasize efficiency, effectiveness and satisfaction [International Standards Organization, 2008], but they relegate metrics such as pleasure and emotion to the sidelines. This study describes that while performance metrics do not show much difference, mental workload (measured via the task-evoked pupillary response) shows a difference with/without support for synchronization (in the Contacts task).

Many devices that are intended to be used in collaboration with other devices are designed independently of one another. In some cases, it appears as if minimal attention has been given during the design process to understand the broader context of use and to situate the device in this context, offering support for the activities that are performed in real use scenarios. When evaluated for usability, many devices are often tested in pristine laboratory settings. Even if tested in real world scenarios, they may not be evaluated together with other interacting devices in the user's work environment. The lack of correlation in this experiment between task metrics and workload measures — despite systemic differences in each individually — stresses the need for conducting holistic usability evaluations of such devices when they act together to fulfill a user's information needs. Tasks are no longer confined to single devices; users are likely to perform part of a task on one device, and part on another. Device designers need to recognize the need for maintaining task context across devices, and pay special attention to the task migration steps that need to be undertaken when two or more devices are used for performing a single task. Systems must also strive to provide adequate support for this critical step, requiring as little user effort as possible.

6.3 Future Work

Practically every research study raises more questions than it answers, as it should. While both my studies have provided a much clearer understanding of the problems facing multi-device PIM than prior studies, there are several questions that would benefit from a more detailed focused study.

6.3.1 Investigating the Applicability of Workload Assessment in PIM Tasks

This study provides evidence to warrant further studies of other forms of non-traditional usability metrics for personal information tasks. One of the two scales used in this experiment was not sensitive enough to different task conditions; future experiments with other subjective workload assessment techniques can yield insights into which, if any, of them are useful in office environments for tasks such as these. It is also interesting to conduct interview studies to determine if users' choices of device are correlated in any way to their subjective assessments of experience and workload measured via these techniques.

6.3.2 Technology Adoption Issues

What are some of the reasons that people buy and use the devices they do? How closely is their long-term use and acceptance of their devices related to the symbiotic relations it forms with their other devices? A longitudinal ethnographic study of these aspects would provide valuable insights into the gap between *good enough* and *insanely great* products on the market.

6.3.3 A Closer Look at Task Migrations

While this experiment provided confirmation that the level of system support for task migration affects both, time taken for migration, and mental workload, this is an area that requires deeper investigation. Task migrations are of several types, may occur between/among several devices, and may occur across varying amounts of time intervals. What are the some of the aspects that may be automated, and what types of system support need to be provided?

6.3.4 Evaluating the Syncables framework

Earlier, I developed the Syncables framework [Tungare et al., 2007] aimed towards bridging some of the issues in task migration: specifically, information migration. I would like to test its effectiveness in multiple tasks in ways similar to this experiment. In addition, any such framework must be easy to develop for; thus, evaluating this software framework from the point of view of software developers would be important as an indicator of developer acceptance.

6.3.5 Measuring Equilibrium in Personal Information Ecosystems

In a previous paper [Pérez-Quiñones et al., 2008], we discussed Personal Information Ecosystems, and the concept of equilibrium in them when information flows seamlessly from one device to another. It is interesting to note that for such a system to be of the utmost benefit to users, it must cause minimal mental workload. Thus, it can be hypothesized that mental workload can serve as a metric of equilibrium in personal information ecosystems. I would like to evaluate this claim by instantiating several versions of device ecosystems and conducting user studies.

Chapter 7

Appendices

7.1 Survey Questionnaire

Your Devices

1. Which of the following devices do you own and use regularly? How many of each type do you have?

	None	One	Two	Three	Four or more
Desktop computer (work)					
Desktop computer (home)					
Laptop computer (either work or home)					
Portable media player (e.g. iPod)					
Cell phone					
Personal Digital Assistant (PDA)					
Treo, Blackberry, iPhone, other multi-function device					
Digital camera					

2. What activities do you usually perform on each of your devices? If a feature exists, but you do not use it, please do not check the box.

Not all choices will apply to all devices. Please check all boxes that apply in your case. Primary home and work computers could be your desktop or laptop computer, as appropriate.

	Desktop (work)	Desktop (home)	Laptop	Cell phone	Portable media player	PDA	Treo, Blackberry, iPhone, or other	Digital camera
Browsing the Web								
Read web email (work)								
Read web email (personal)								
Download email (work)								
Download email (personal)								
Instant messaging (work)								
Instant messaging (personal)								
Send / receive SMS								
Address book / contacts (work)								
Address book / contacts (personal)								
Calendar (work)								
Calendar (personal)								
Read or edit documents (work)								
Read or edit documents (personal)								
To-do notes								
Making phone calls								
Playing music								
Watching videos								
Taking photos								
Storing, viewing or managing photos								

3. If you own a multi-function device (e.g. Treo, Blackberry, iPhone), has the presence of this device led you to abandon any other device (e.g. devices that previously performed each individual function.) If yes, please tell us more about the multi-function device as well as the others it replaced.

You can enter as many lines as you want; don't let the size of the box limit your response.

Using Multiple Devices Together

4. Which of the above devices do you frequently operate nearly at the same time? E.g. at your office, you might use your PDA and your desktop simultaneously. On the go, you might always carry your iPod and phone. In each row below, select the devices that are used together in a group.

Feel free to use as many rows as you need and leave the rest blank.

	Work desktop	Home desktop	Laptop	PDA	Cell phone	Portable media player	Treo, Blackberry, iPhone	Other multi-function device
Group 1	☐	☐	☐	☐	☐	☐	☐	☐
Group 2	☐	☐	☐	☐	☐	☐	☐	☐
Group 3	☐	☐	☐	☐	☐	☐	☐	☐
Group 4	☐	☐	☐	☐	☐	☐	☐	☐
Group 5	☐	☐	☐	☐	☐	☐	☐	☐

5. Between which pairs of devices do you usually copy or synchronize data?

Feel free to use as many rows as you need and leave the rest blank.

	First Device	Direction	Second Device	Type of data
Pair 1	Work desktop	Synchronizes both ways	Home desktop	Documents
Pair 2	-- Select --	-- Select --	-- Select --	-- Select --
Pair 3	-- Select --	-- Select --	-- Select --	-- Select --
Pair 4	-- Select --	-- Select --	-- Select --	-- Select --
Pair 5	-- Select --	-- Select --	-- Select --	-- Select --
Pair 6	-- Select --	-- Select --	-- Select --	-- Select --
Pair 7	-- Select --	-- Select --	-- Select --	-- Select --
Pair 8	-- Select --	-- Select --	-- Select --	-- Select --
Pair 9	-- Select --	-- Select --	-- Select --	-- Select --
Pair 10	-- Select --	-- Select --	-- Select --	-- Select --

6. Data synchronization horror stories: Syncing data sometimes has its own pitfalls. Have you ever been victim to a situation where synchronization failed to live up to your expectations, either due to system errors, or because of forgetting to do it, etc.? If you have a story, please share with us. When was the last time such an incident happened to you?

You can enter as many lines as you want; don't let the size of the box limit your response.

Buying a New Device

7. Please indicate your agreement with the statement below for each factor in the left column:

"Factor X is the single most important factor to me when buying a new device."

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
Feature richness	○	○	○	○	○
Price	○	○	○	○	○
Ease of use	○	○	○	○	○
"Hipness"	○	○	○	○	○
A good fit with existing devices	○	○	○	○	○
Manufacturer/brand, etc.	○	○	○	○	○

8. Overall, how satisfied have you been with the last device you purchased?

○ Very satisfied
○ Satisfied
○ Neither satisfied nor dissatisfied
○ Dissatisfied
○ Very dissatisfied

9. If you ran into any problems using your new device with your existing devices and data, please describe them. Feel free to leave blank if you were entirely satisfied with how your new device integrated into your life.

You can enter as many lines as you want; don't let the size of the box limit your response.

10. Have you ever encountered a situation where one of your devices stopped functioning, or was otherwise unusable for its normal function?

Feel free to use as many rows as you need and leave the rest blank.

	What device?	Did you lose data?	Was it a hardware or software issue?	Were you able to restore your data from a backup copy?	How soon did you get a device to replace the failed one?	How long did it take for the new device to completely replace the function of the failed device?
Failed Device	-- Select --	-- Select --	-- Select --	-- Select --	-- Select --	-- Select --

1						
Failed Device 2	-- Select --	-- Select --	-- Select --	-- Select --	-- Select --	-- Select --
Failed Device 3	-- Select --	-- Select --	-- Select --	-- Select --	-- Select --	-- Select --
Failed Device 4	-- Select --	-- Select --	-- Select --	-- Select --	-- Select --	-- Select --
Failed Device 5	-- Select --	-- Select --	-- Select --	-- Select --	-- Select --	-- Select --

About you

11. Are you male or female?

◯ Male ◯ Female

12. Which of the following age groups do you belong into?

-- Select --

13. What is the highest level of education you have completed?

-- Select --

14. Do you consider yourself an information worker (or a knowledge worker)?

◯ Yes, full-time ◯ Yes, part-time ◯ No ◯ Not sure

15. Who manages your calendar appointments? Please check all boxes that apply in your case.

☐ You ☐ Your assistant ☐ Your spouse ☐ Your parent ☐ Other (please specify) ________

16. Which of the following does your primary work activity involve? Please check all boxes that apply in your case.

☐ Working at a desk
☐ Communicating with people
☐ Conducting research
☐ Attending classes
☐ Traveling locally (roughly within the same city, town, or metropolitan area)
☐ Traveling between local offices (but no airline travel)
☐ Airline travel
☐ Other (please specify)

____

17. What is your primary mode of transport for commuting to your workplace?

◯ None, I telecommute
◯ Walk
◯ Use a bicycle
◯ Drive
◯ Carpool
◯ By train
◯ By bus

18. How long is your one-way commute each day?

◯ I telecommute
◯ Less than 10 minutes
◯ 10-20 minutes
◯ 20-40 minutes
◯ 40 minutes to an hour

7.2 IRB Approval for Survey

VirginiaTech

Office of Research Compliance
Carmen T. Green, IRB Administrator
2000 Kraft Drive, Suite 2000 (0497)
Blacksburg, Virginia 24061
540/231-4358 Fax 540/231-0959
e-mail ctgreen@vt.edu
www.irb.vt.edu
FWA00000572(expires 1/20/2010)
IRB # is IRB00000667

DATE: August 7, 2007

MEMORANDUM

TO: Manuel A. Perez-Quinones
Manas Tungare

FROM: Carmen Green

SUBJECT: **IRB Exempt Approval:** "Understanding the Evolution of Users' Personal Information Management Practices" , IRB # 07-388

I have reviewed your request to the IRB for exemption for the above referenced project. I concur that the research falls within the exempt status. Approval is granted effective as of August 7, 2007.

As an investigator of human subjects, your responsibilities include the following:

1. Report promptly proposed changes in previously approved human subject research activities to the IRB, including changes to your study forms, procedures and investigators, regardless of how minor. The proposed changes must not be initiated without IRB review and approval, except where necessary to eliminate apparent immediate hazards to the subjects.

2. Report promptly to the IRB any injuries or other unanticipated or adverse events involving risks or harms to human research subjects or others.

cc: File

Invent the Future

VIRGINIA POLYTECHNIC INSTITUTE UNIVERSITY AND STATE UNIVERSITY

An equal opportunity, affirmative action institution

7.3 IRB Requirements for Experiments

7.3.1 Approval Letter

VirginiaTech

Office of Research Compliance
Institutional Review Board
2000 Kraft Drive, Suite 2000 (0497)
Blacksburg, Virginia 24061
540/231-4991 Fax 540/231-0959
e-mail moored@vt.edu
www.irb.vt.edu

FWA00000572(expires 1/20/2010)
IRB # is IRB00000667

DATE: October 27, 2008

MEMORANDUM

TO: Manuel A. Perez-Quinones
Manas Tungare

Approval date: 10/27/2008
Continuing Review Due Date:10/12/2009
Expiration Date: 10/26/2009

FROM: David M. Moore

SUBJECT: **IRB Expedited Approval:** "Understanding Users' Personal Information Management Practices Across Devices" , IRB # 08-652

This memo is regarding the above-mentioned protocol. The proposed research is eligible for expedited review according to the specifications authorized by 45 CFR 46.110 and 21 CFR 56.110. As Chair of the Virginia Tech Institutional Review Board, I have granted approval to the study for a period of 12 months, effective October 27, 2008.

As an investigator of human subjects, your responsibilities include the following:

1. Report promptly proposed changes in previously approved human subject research activities to the IRB, including changes to your study forms, procedures and investigators, regardless of how minor. The proposed changes must not be initiated without IRB review and approval, except where necessary to eliminate apparent immediate hazards to the subjects.
2. Report promptly to the IRB any injuries or other unanticipated or adverse events involving risks or harms to human research subjects or others.
3. Report promptly to the IRB of the study's closing (i.e., data collecting and data analysis complete at Virginia Tech). If the study is to continue past the expiration date (listed above), investigators must submit a request for continuing review prior to the continuing review due date (listed above). It is the researcher's responsibility to obtain re-approval from the IRB before the study's expiration date.
4. If re-approval is not obtained (unless the study has been reported to the IRB as closed) prior to the expiration date, all activities involving human subjects and data analysis must cease immediately, except where necessary to eliminate apparent immediate hazards to the subjects.

Important:
If you are conducting **federally funded non-exempt research**, please send the applicable OSP/grant proposal to the IRB office, once available. OSP funds may not be released until the IRB has compared and found consistent the proposal and related IRB applicaton.

cc: File

Invent the Future

VIRGINIA POLYTECHNIC INSTITUTE UNIVERSITY AND STATE UNIVERSITY

An equal opportunity, affirmative action institution

7.3.2 IRB-Approved Consent Form

VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY

Informed Consent for Participants in Research Projects Involving Human Subjects

Understanding Users' Personal Information Management Practices

Investigator(s): Dr. Manuel Pérez-Quiñones, Manas Tungare

I. Purpose of this Research/Project

As part of our research about how users use multiple devices to manage their personal information, we are conducting a set of experiments to gain a deeper understanding of the issues involved. We expect to recruit about 20 participants for this research. Our participants are knowledge workers who regularly use information devices such as laptop computers and cell phones. Thank you for participating in our research.

II. Procedures

This experiment will be carried out in two sessions of 1 hr each, separated by a period of two weeks. During each session, we will request you to perform several experimental tasks that involve laptop computers, desktop computers, cell phones, and personal digital assistants (PDAs). The tasks involve everyday activities such as copying information, making phone calls and editing files. While you perform these tasks, we wish to examine how you perform them, with the assistance of equipment such as eye trackers. As part of the experiment, you will be requested to wear head-mounted eye tracking equipment. If you do not feel comfortable using this equipment, you are free to opt out of the experiment at any time. At the end of each task, we will also ask you to fill a quick questionnaire about your opinion of the task you just performed. At the end of each experimental session, we will conduct a short interview about your information management practices, which is expected to last no more than 20 minutes.

We are not evaluating you or judging the practices you employ; rather, we plan to collect and analyze this information from several participants, identify some of the common areas where there is a gap between the ideal situation and current practices, and possibly make recommendations regarding today's tools.

The eye tracking software captures a video feed of what you look at. We will not be using any additional cameras to record your actions. Interviews conducted at the conclusion of the experimental tasks will be recorded as audio to be transcribed and analyzed later.

III. Risks

There are no more than minimal risks involved during this experiment. The tasks we request you to perform are very likely common tasks that you perform each day.

IV. Benefits

There are no direct benefits to you for participating in this research — no promise or guarantee of benefits have been made to encourage you to participate. Indirect benefits may include the development of better tools to manage your personal information that may result as recommendations from this study. You are welcome to contact us in due time if you are interested in the findings of this research.

V. Extent of Anonymity and Confidentiality

The data we collect during this interview will be anonymized using study codes (i.e., your responses will be identified only as P_1, P_2, etc. (P = Participant.)) Eye tracking data, questionnaire responses and interview data will all be anonymized prior to analysis. After interviews are transcribed, the original recordings will be de-

VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY

stroyed. You may opt out of this interview at any point during the process. If you choose to opt out, all the data recorded or noted down during this session will be immediately destroyed.

The data collected from these interviews may be reported by the researchers in academic conferences, journals, and as part of students' dissertations. In no such publication will any identifying information be included. Anonymized reporting may refer to participants with their coded identifiers. It is possible that the Institutional Review Board (IRB) may view this study's collected data for auditing purposes. The IRB is responsible for the oversight of the protection of human subjects involved in research.

VI. Compensation

In return for your time performing this experiment, we will provide gift certificates worth $10 for your participation. If you choose to participate in a single session only, the compensation will be pro-rated based on the time spent.

VII. Freedom to Withdraw

You are free to withdraw from this study at any time without penalty. You are free to refuse to answer any questions without penalty.

VIII. Subject's Responsibilities

You are responsible for abiding by the terms of any non-disclosure agreements that you may be a party to at the time this interview is conducted.

IX. Subject's Permission

I have read the Consent Form and conditions of this project. I have had all my questions answered. I hereby acknowledge the above and give my voluntary consent:

____ Date: **2008**- ____-____.
Participant's signature

Should I have any pertinent questions about this research or its conduct, and research subjects' rights, and whom to contact in the event of a research-related injury to the subject, I may contact:

Investigator: **Manas Tungare**
Telephone: (650)862-3627
Email: manas@vt.edu

Investigator and Faculty Advisor: **Dr. Manuel Pérez-Quiñones**
Email: perez@cs.vt.edu

Chair, Virginia Tech Institutional Review Board for the Protection of Human Subjects: **David M. Moore**
Telephone: (540)231-4991
Email: moored@vt.edu

Virginia Tech Institutional Review Board: Project No. 08-652
Approved October 27, 2008 to October 26, 2009

7.4 Experimenter's Script for Study 2

Script for Study on Personal Information Management

Thank you for participating in this experiment. My name is Manas and I will be assisting you today. If you have any questions about any part of today's experiment, please feel free to ask me.

We are researching how people use multiple devices when managing their personal information. This includes data such as files, calendar events, and contact information that is managed on devices such as multiple computers, cell phones, PDAs, etc.

Before we proceed, I would like to know whether we have your informed consent to proceed with this experiment. Please take a few minutes to read this consent form and if you agree, please sign it at the bottom.

Give consent form to participant and wait for response. If not signed, say thanks and do not proceed.

Thanks for the consent form! We'd like to know a few demographics about you: here is a questionnaire.

Give them a few minutes to enter the responses to the pre-questionnaire.

We will now start by performing a few common tasks to familiarize you with the software. Many of these are simple office tasks that you might have performed several times in the past. Further instructions will be available to you on the large screen, one at a time. Please make sure to let the experimenter know when you have finished each step, so we may advance to the next instruction.

Do Training Tasks
Do NASA TLX

That brings us to the end of the training tasks, so we will proceed to the experimental tasks. There are three experimental tasks, each of which usually takes about 8 to 10 minutes. We will be recording the time taken for each task, so please try to complete each as quickly as possible.

Before we do that, let's put on the eye tracking device and calibrate it.

Assist them in wearing the eye tracker.

- ***Put eye tracker on their head***
- ***Confirm that you see their pupil size***
- ***Start the video tape***
- ***Start the CSV file logging***
- ***Do the calibration routine: for pupil and for gaze***
- ***Confirm that data is being saved correctly in (C:\Program Files\...)***

Once the experimental tasks are done,

Let us now remove the eye tracker.

Assist them in removing the eye tracker.

- ***Stop the recording.***
- ***Stop the CSV logging.***

- ***Turn off the eye tracker.***
- ***Take eye tracker off.***
- ***Place it out of their way.***

Thanks for participating in our experiment! Here is a token of appreciation for your trouble.

Give them the gift certificate.

This was session 1 of 2. Let's confirm the date and time of the second session. The second session will be shorter than the first because we will no longer need to do the same training tasks in the beginning.

Look at your calendar, confirm their schedules.

Say thanks, and escort them out of the room/building.

7.5 Demographic Questionnaire

Participant Code:	Date:	Treatment:	Session:

Pre-Questionnaire

1. Are you male or female?
   ☐ Male
   ☐ Female

2. Which of the following age groups do you belong to? (select only one.)
   ☐ Less than 18
   ☐ 18-21 years
   ☐ 22-25 years
   ☐ 26-30 years
   ☐ 31-35 years
   ☐ 36-40 years
   ☐ 41-50 years
   ☐ 51-58 years
   ☐ More than 58

3. What is the highest level of education you have completed?
   ☐ Middle school
   ☐ High school
   ☐ Bachelor's degree
   ☐ Master's degree
   ☐ Doctoral degree

4. Do you consider yourself an information worker (or a knowledge worker)? (If your primary work activity involves working with computers, you can consider yourself to be a knowledge worker.)
   ☐ Yes, full-time
   ☐ Yes, part-time
   ☐ No
   ☐ Not sure

5. How much travel does your regular work activity involve?

   None   Travel within the City/Town   Travel to other cities/towns   International travel

6. Do you engage in travel infrequently, for example, conferences or remote meetings?

   None   Travel within the City/Town   Travel to other cities/towns   International travel

7. Tell us about occasions where your usual way of managing information does not work any more. E.g. while on extended travel. (We will ask you more details about this during a follow-up interview at the end of the second session of our experiment.)

8. What is your primary mode of transport for commuting to your workplace?
   - ☐ None, I telecommute
   - ☐ Walk
   - ☐ Use a bicycle
   - ☐ Drive
   - ☐ Carpool
   - ☐ By train
   - ☐ By bus
   - ☐ other:

9. How long is your one-way commute each day?
   - ☐ I telecommute
   - ☐ Less than 10 minutes
   - ☐ 10-20 minutes
   - ☐ 20-40 minutes
   - ☐ 40 minutes to an hour
   - ☐ Between 1 and 2 hours
   - ☐ More than 2 hours

7.6 Dimensions of the NASA TLX scale

A copy of this was provided to each participant to assist in their subjective evaluations.

Sub Scale	End Points	Description
Mental Demand	Low/High	How much mental and perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searching, etc.) Was the task easy or demanding, simple or complex, exacting or forgiving?
Physical Demand	Low/High	How much physical activity was required (e.g. pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
Temporal Demand	Low/High	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
Performance	Good/Poor	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
Effort	Low/High	How hard did you have to work (mentally and physically) to accomplish your level of performance?
Frustration Level	Low/High	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

7.7 The NASA TLX Scale

The NASA TLX scale is described in [Hart and Staveland, 1988].

Participant Code: Date: Treatment: Session:

Mental Demand

How mentally demanding was the task?

Very low — Very high

Physical Demand

How physically demanding was the task?

Very low — Very high

Temporal Demand

How hurried or rushed was the pace of the task?

Very low — Very high

Performance

How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort

How hard did you have to work to accomplish your level of performance?

Very low — Very high

Frustration

How insecure, discouraged, irritated, stressed and annoyed were you?

Very low — Very high

PLEASE TURN PAGE OVER

Weights

In each pair of factors below, which of the two do you think is more important for the task that you just performed?

Place an "X" mark next to the one you think is more important than the other.

Mental Demand	☐	☐	Physical Demand
Physical Demand	☐	☐	Temporal Demand
Effort	☐	☐	Performance
Frustration Level	☐	☐	Performance
Mental Demand	☐	☐	Effort
Effort	☐	☐	Physical Demand
Frustration Level	☐	☐	Mental Demand
Mental Demand	☐	☐	Temporal Demand
Effort	☐	☐	Frustration Level
Physical Demand	☐	☐	Performance
Temporal Demand	☐	☐	Effort
Mental Demand	☐	☐	Performance
Physical Demand	☐	☐	Frustration Level
Performance	☐	☐	Temporal Demand
Temporal Demand	☐	☐	Frustration Level

7.8 Participant Instructions for Tasks

7.8.1 Files Task, using USB/Email

Task 1: Making Changes to Files (Files L0)

Story

You are a consultant who works with several different clients on multiple projects. For convenience and ease of use, you work on a laptop when mobile, and a desktop computer at your office. At the start of this experiment, we will show you the simulated **office desk**, and the simulated **client location**. You will be required to move between these locations. (For the purpose of this experiment, we will substitute a second desktop computer instead of a laptop. We will not be using the mobility features of the laptop, so you can perform the tasks exactly as you would on a laptop computer.)

On-screen Instructions

Instructions from clients will be delivered to you on-screen. When each instruction is received, you should act on it and make the changes requested in it.

Types of Files

Each instruction will clearly identify the specific client who made the request, and specific file that you will need to edit. There are several files into which such information may go. All these files are stored on the hard disk of the laptop and desktop computer. For this experiment, we will use three types: (1) **Notes** or **Lists** (plain text files), (2) **Presentations** or **Slides**, and (3) **Spreadsheets**.

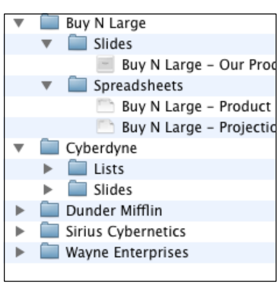
Moving Between Your Desktop and Laptop

Some instructions will suggest a change in location. When you receive such an instruction, you need to finish your pending work at the current location (either your office desk or client location.) It is also your responsibility to ensure that you **carry updated copies of files** when moving between locations, because subsequent tasks will require you to make changes to the same files that you modified earlier. You may then proceed to move to the location indicated on that index card.

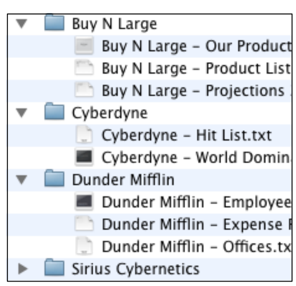
There are several companies and files on your disk; to copy them from one location to another will take about 15-20 minutes.

Which of the following file and folder structures resembles your hard disk closely?

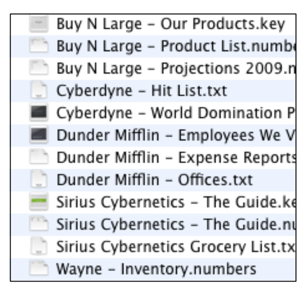
Because you might be accustomed to a certain way of managing your files, we would like to present you with a file system that looks closely like your own hard disk.



Deeply Nested



Moderately Nested



Flat Hierarchy

Does it look more like (1), (2) or (3)? In (1), there are several folders and subfolders, with only a few files inside each subfolder. In (2), there are few folders, and all files are located directly inside them. In (3), there is only the top-level folder, and all files stored directly beneath it. At this point, please stop reading, and tell the experimenter which one you picked.

Companies

Your clients include the following fictitious companies: (the names in bold indicate the short forms used in the task descriptions): **Buy N Large** Corporation, **Cyberdyne** Systems, **Dunder Mifflin** Paper Company, **Sirius** Cybernetics Corporation, **Wayne** Enterprises.

Objective

In the end, your objective is to be sure that all the files on your laptop as well as on the desktop are kept updated with all the changes informed to you so far.

Devices and Resources

For this task, you will be provided a USB drive. You may use the drive to copy files between your two machines. The USB drive can be inserted in the USB socket provided at the back of the keyboard on both machines. If you have any questions about this, please feel free to ask the experimenter.

You will also have access to an email account that is set up on both machines. You may choose to use it to transfer your files by emailing it to yourself. A link to the web-based email provider is present on the Desktop of both computers. Use the email address **participant@manastungare.com** with the password **pimstudy**.

7.8.2 Files Task, using Network Drive

Task 1: Making Changes to Files (Files L1)

Story

You are a consultant who works with several different clients on multiple projects. For convenience and ease of use, you work on a laptop when mobile, and a desktop computer at your office. At the start of this experiment, we will show you the simulated **office desk**, and the simulated **client location**. You will be required to move between these locations. (For the purpose of this experiment, we will substitute a second desktop computer instead of a laptop. We will not be using the mobility features of the laptop, so you can perform the tasks exactly as you would on a laptop computer.)

On-screen Instructions

Instructions from clients will be delivered to you on-screen. When each instruction is received, you should act on it and make the changes requested in it.

Types of Files

Each instruction will clearly identify the specific client who made the request, and specific file that you will need to edit. There are several files into which such information may go. All these files are stored on the hard disk of the laptop and desktop computer. For this experiment, we will use three types: (1) **Notes** or **Lists** (plain text files), (2) **Presentations** or **Slides**, and (3) **Spreadsheets**.

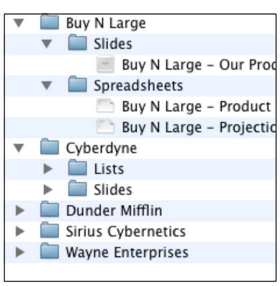
Moving Between Your Desktop and Laptop

Some instructions will suggest a change in location. When you receive such an instruction, you need to finish your pending work at the current location (either your office desk or client location.) It is also your responsibility to ensure that you **carry updated copies of files** when moving between locations, because subsequent tasks will require you to make changes to the same files that you modified earlier. You may then proceed to move to the location indicated on that index card.

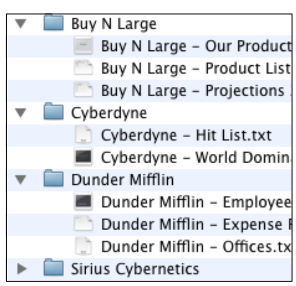
There are several companies and files on your disk; to copy them from one location to another will take about 15-20 minutes.

Which of the following file and folder structures resembles your hard disk closely?

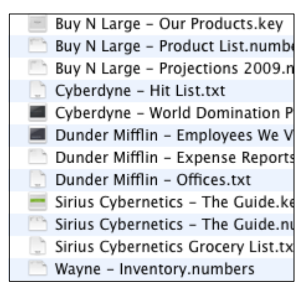
Because you might be accustomed to a certain way of managing your files, we would like to present you with a file system that looks closely like your own hard disk.



Deeply Nested



Moderately Nested



Flat Hierarchy

Does it look more like (1), (2) or (3)? In (1), there are several folders and subfolders, with only a few files inside each subfolder. In (2), there are few folders, and all files are located directly inside them. In (3), there is only the top-level folder, and all files stored directly beneath it. At this point, please stop reading, and tell the experimenter which one you picked.

Companies

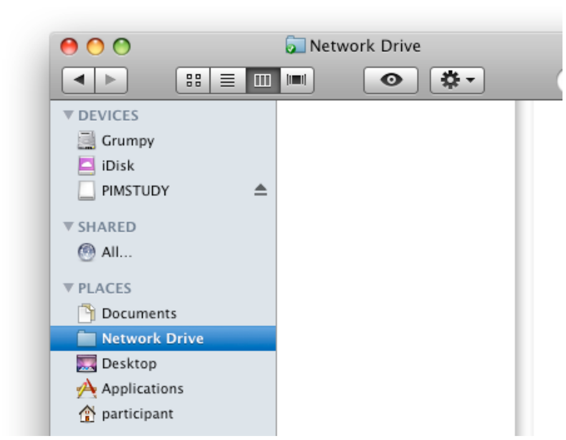
Your clients include the following fictitious companies: (the names in bold indicate the short forms used in the task descriptions): **Buy N Large** Corporation, **Cyberdyne** Systems, **Dunder Mifflin** Paper Company, **Sirius** Cybernetics Corporation, **Wayne** Enterprises.

Objective

In the end, your objective is to be sure that all the files on your laptop as well as on the desktop are kept updated with all the changes informed to you so far.

Devices and Resources

For this task, you will be provided access to a Network Drive. A Network Drive is a place to store your files such that they can be accessed from any machine, not just the one that you use to put your files there. To access the network drive from either of your machines, please click the icon labeled "Network Drive" in the Finder or in the Dock.



7.8.3 Calendar Task, using Paper Calendars

Task 2: Managing Calendars (Calendar L0)

Story

You work at an office with several colleagues. Part of your job is to meet regularly with your colleagues, and manager. Since work and family are two distinct parts of your life, you have chosen to maintain separate calendars for the two types of events. Both calendars are exactly similar, except for the type of events you write in them.

On-Screen Instructions

As new meetings or parties are scheduled, the experimenter will inform you about these events via instructions on the screen. Each instruction will contain details related to an event: who wants to schedule it, what it is for, when it is, and where it will be held. Some instructions may also include a question posed to you: as soon as you have the answer available, please let the experimenter know.

Tentative Meetings

Sometimes, a meeting might be scheduled tentatively because it involves several people who need to confirm their availability before a final schedule can be decided. The on-screen instructions will indicate whether a meeting request is final or tentative. You should attempt your best to make sure that you can accommodate most meetings without conflicts.

Devices and Resources

For this task, you will be provided two paper calendars: Home and Work. Personal events are recorded on the Home calendar, while work-related events are recorded on the Work calendar.

7.8.4 Calendar Task, using Online Calendar System

Task 2: Managing Calendars (Calendar L1)

Story

You work at an office with several colleagues. Part of your job is to meet regularly with your colleagues, and manager. Since work and family are two distinct parts of your life, you have chosen to maintain separate calendars for the two types of events. Both calendars are exactly similar, except for the type of events you write in them.

On-Screen Instructions

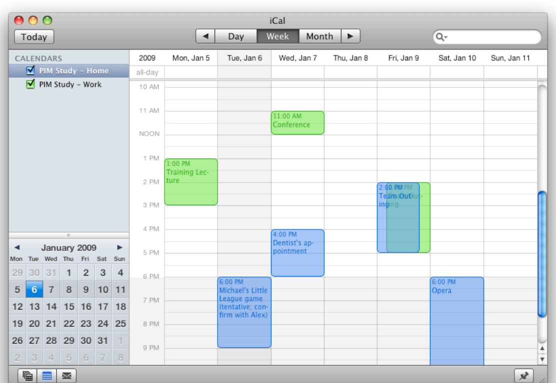
As new meetings or parties are scheduled, the experimenter will inform you about these events via instructions on the screen. Each instruction will contain details related to an event: who wants to schedule it, what it is for, when it is, and where it will be held. Some instructions may also include a question posed to you: as soon as you have the answer available, please let the experimenter know.

Tentative Meetings

Sometimes, a meeting might be scheduled tentatively because it involves several people who need to confirm their availability before a final schedule can be decided. The on-screen instructions will indicate whether a meeting request is final or tentative. You should attempt your best to make sure that you can accommodate most meetings without conflicts.

Devices and Resources

For this task, you will be provided an online calendaring tool, Apple iCal. Within the program, you will find two calendars, Home and Work. The home calendar contains personal events, while the work calendar contains work-related events.



7.8.5 Contacts Task, without Synchronization Software

Task 3: Managing Contact Information (Phone L0)

Story

You are attending a conference/business meeting, and meet several new people there. You also meet a lot of old professional colleagues who you have been out of touch with. A lot of people give you their business cards – many who you are meeting for the first time, and many whose phone numbers have been updated since you last met them. Your job is to update the contact list on your cell phone and your laptop, from the business cards that you receive from your colleagues. (For the purpose of this experiment, we will use a desktop computer in place of a laptop.)

You will meet colleagues in two contexts:

- During a session, your laptop is powered on, while your cell phone has been turned off. You are not allowed to use your cell phone when the task specifies as such.
- Between sessions, you will meet colleagues in the corridors where your laptop is not available for use, but your cell phone is handy. You are not allowed to use your laptop when the task specifies as such.

A series of events will occur, delivered to you via on-screen instructions. Some instructions will also require you to contact some of these colleagues via phone or email. You should find out the relevant information asked in the on-screen instructions, and inform the experimenter when you have the answers.

Important!

For the tasks related to Contacts, do not worry about whether a particular email address or phone number is a **work** phone or **home** phone. In this experiment, we will not differentiate between these, so ignore these even if the software makes a provision for entering it.

Devices and Resources

Unfortunately, your computer and cell phone cannot be automatically synchronized, so you will need to obtain the information you need from both these sources, as required by the specific lookup task.

7.8.6 Contacts Task, with Synchronization Software

Task 3: Managing Contact Information (Phone L1)

Story

You are attending a conference/business meeting, and meet several new people there. You also meet a lot of old professional colleagues who you have been out of touch with. A lot of people give you their business cards – many who you are meeting for the first time, and many whose phone numbers have been updated since you last met them. Your job is to update the contact list on your cell phone and your laptop, from the business cards that you receive from your colleagues. (For the purpose of this experiment, we will use a desktop computer in place of a laptop.)

You will meet colleagues in two contexts:

- During a session, your laptop is powered on, while your cell phone has been turned off. You are not allowed to use your cell phone when the task specifies as such.
- Between sessions, you will meet colleagues in the corridors where your laptop is not available for use, but your cell phone is handy. You are not allowed to use your laptop when the task specifies as such.

A series of events will occur, delivered to you via on-screen instructions. Some instructions will also require you to contact some of these colleagues via phone or email. You should find out the relevant information asked in the on-screen instructions, and inform the experimenter when you have the answers.

Important!

For the tasks related to Contacts, do not worry about whether a particular email address or phone number is a **work** phone or **home** phone. In this experiment, we will not differentiate between these, so ignore these even if the software makes a provision for entering it.

Devices and Resources

The contacts on your laptop and cell phone can easily be synchronized by attaching a cable and pressing a button. This will assist you in answering some of the questions posed at the end of the task. The experimenter will demonstrate to you the details of synchronizing these devices.

7.9 Task Instructions

This appendix lists the set of tasks and instructions presented to participants for each step of the task. It must be noted that although the steps are presented here as a single list, this was not how they were administered during the experiment. Section §3.6.7 describes in detail the procedure that was used to administer instructions one at a time to participants.

7.9.1 Familiarization Task Instructions

The familiarization procedure, including the videos created, are presented in section §3.6.2. This appendix provides a detailed list of the specific familiarization tasks administered to participants during the experiment.

0. You will now perform a set of simple office tasks.

1. Using the spreadsheet program, add 3 columns to a new spreadsheet:
   Student Name, Registration Number, and Year.

2. Add the details of two students as follows.

   (a) John Doe, 154-974-2546, 2008

   (b) Mona Lisa, 874-376-3467, 2007

3. Using the presentation program, add two slides to a new presentation.

Title: My First Slide

Title: My Second Slide • First Bullet • Second Bullet

4. In the calendar program, please add the following event.
   Presidential Inauguration
   Date: January 20, 2009.
   Time: 12:00 noon to 2:00 pm.

5. In the address book manager, please modify the contact details as follows:
   Sheldon Cooper
   From Work Phone: 626-555-1234
   To Work Phone: 626-974-3468

6. On the phone provided to you, please modify the contact details as follows:
   Howard Wolowitz
   From Work Phone: 310-459-2434
   To Work Phone: 310-345-7462

7. We're done with the training tasks.
   The experimenter will now outfit you with an eye tracker device.

7.9.2 Files Task Instructions

0. Please read the instructions related to this task, provided to you by the experimenter.

1. You're now at your office. All your files are in the Documents folder. Please proceed to your desk and settle down. Let me know when you're ready to begin.

2. *Wayne Enterprises* sends you an email asking to add *car tires* to *Shopping List*.

3. *Buy N Large* asks you to add a picture of *Cereal* to *Slide 3* of the presentation, *Our Products*.

4. *Dunder Mifflin* just called, they want you to add *Holiday Decorations* to the *Expense Reports* spreadsheet.

5. You need to visit a client, **Dunder Mifflin's Office**. Make sure that all your files will be available on your laptop before you get there.

6. *Sirius* requests you to change the spreadsheet entry for *Earth* in *The Guide* from *"Harmless"* to *"Mostly Harmless"*.

7. *Buy N Large* is not confident about next years Projections. They suggest changing profit outlook from $200B to $150B for First Quarter of 2009 (Q1 2009)

8. *Sirius Cybernetics* suggests adding *English Tea* to *Grocery List*.

9. *Dunder Mifflin* is closing its *Scranton, PA* office. You need to remove it from the *Offices* list.

10. Please return to your own office now. This is the last task, so close all files and please make sure that all the modifications you made (both, at your office and at Dunder Mifflin's office) are now all available on your desktop in the **original directory**.

11. That was the last step in the Files task. The experimenter will now give you a questionnaire to understand how you felt during this task.

7.9.3 Calendar Task Instructions

0. Please read the instructions related to this task, provided to you by the experimenter.

1. Today is January 5, 2009.

2. Your colleague Peter just called, he would like to schedule an hour-long meeting with you today. He is free between noon and 4:00p. All meetings are held between 9:00am and 5:00pm. Propose a time and schedule a meeting.
What time did you schedule?

3. Your spouse Alex sends you a copy of the Opera tickets s/he bought for Saturday night, January 10. The event is from 7:00 to 9:00, but considering traffic, you will need at least an hour to get there and back. Schedule this in your calendar.
What time did you schedule?

4. Your boss, Alice, would like to meet you and three other colleagues, Bob, Carol and Dave on Wednesday for about 2 hours. She has sent a common email to all of you asking you to pick a time that works for everyone (except from 9:00 to 11:00). The meeting must happen on Wednesday.
Make a tentative entry in your calendar. What time did you schedule?

5. Today is January 6, 2009.

6. Bob replies that he is meeting a client for lunch on Wednesday that will last from 11:30a to 1:30p. Make changes to your tentative entry as appropriate and propose a meeting time that works for everyone who has replied so far.
What time did you propose?

7. A friend, Douglas, has left you a voicemail asking if you and your spouse Alex can join him and his wife for dinner on Saturday around 6:00pm.
Can you?

8. Your team just bagged a new contract, and your office buddies are going out for drinks tonight. They'd like to know if you can join them, say at 7:00pm.
Check your appointments for today and tell them what you think.

9. Since todays Little League game was marked as tentative, you decide to confirm with Alex. Turns out it has been moved to Thursday, same time. Also, you need to drive him there, and it takes half an hour.
Note this in your calendar. What time did you schedule?

10. With this new information, call your colleagues to inform them whether or not you can join them for drinks.
    What did you decide?

11. Today is January 7, 2009.

12. The meeting time you proposed has been accepted by everyone.
    What time is your meeting with your boss, Alice, Bob, Carol and Dave today?

13. *What time must you go to the dentist today?*

14. The dentist says you need to come back on Friday so he can do some more work on your teeth.
    What are the possible times you can go?

15. Today is January 8, 2009.

16. *What is the latest time you can get home at, and still not miss any responsibilities?*

17. That was the last step in the Calendar task. The experimenter will now give you a questionnaire to understand how you felt during this task.

7.9.4 Contacts Task Instructions

0. Please read the instructions related to this task, provided to you by the experimenter.

1. During the first session of the conference, you are working on your laptop, and meet Dr. John Smith. He is interested in your research and hands you his business card to stay in touch.

John Smith, Ph.D. Bradbury University Savannah, GA Home: 912-336-8637 smith@bradbury.edu

   Add it to your laptop.

2. You have just shut down your laptop and stepped out for lunch. During the lunch hour, you see an old friend, Anand Narayan, who you know from past collaborations. You decide to meet up for dinner tomorrow evening, and — just to be sure — you decide to confirm that his phone number is still the same.
   What is his phone number?

3. Anand tells you that his cell phone number has changed, and it is now 312-867-3184.
   Update the number on your phone.

4. During the afternoon session, you asked a question to the presenter related to your current research. After the talk, a person approaches you and says that they would love for you to send them more info about your research.

Rodrigo Diaz Post-Doctoral Associate rjdiaz37@uprm.edu 787-376-6673

   Note down their name and email address on your phone.

5. During the evening session, you're pleasantly surprised to see one of your ex-students in the seat next to you. You invite them to join you for dinner with Anand the next evening. They give you a phone number on a sticky note.

Peter Jackson 408-232-4583

   Make this entry on your laptop.

6. At your hotel that night, you decide to email Rodrigo Diaz and Prof. Smith about your research.
   What email addresses will you send it to?

7. The next day morning, you decide to make plans for dinner, and call Anand and Peter.
   What numbers will you call for each one of them?

8. That was the last step in the Contacts task. The experimenter will now give you a questionnaire to understand how you felt during this task.

7.10 Analysis Scripts

7.10.1 PupilSmoother.R

```
# This script runs a 4th order Savitzky-Golay filter of size 151 to smooth raw pupil data.

source("Common.R");
require(signal);

# Open files, target only the task-specific ones, ignore the common *.pupil file.
pupilFiles = list.files(path = "../Generated/", pattern = " [A-Za-z]*\\.pupil$", full.names = TRUE);

for (pupil.file in pupilFiles) {
  print(paste("Reading: ", pupil.file, sep = ""));
  pupil.data = read.table(pupil.file, header = TRUE);

  # The actual smoothing step. Do it on a single column, PupilR.
  smoothed.data = sav.gol(pupil.data$PupilR, 151);

  # Merge that column back into the rest of the data frame.
  for (reading in 1:length(pupil.data$PupilR)) {
    if (!is.na(smoothed.data[reading])) {
      pupil.data$PupilR[reading] = round(smoothed.data[reading], digits=3);
    }
  }

  # Write output to *.pupil.smooth file.
  pupil.outputFile = sub(".pupil", ".pupil.smooth", pupil.file);
  print(paste("Writing: ", pupil.outputFile));
  write.table(pupil.data, file = pupil.outputFile, row.names = FALSE, quote = FALSE);
}
```

7.10.2 PupilAdjuster.R

```
# This script calculates a baseline value for pupil radius from the first 5 seconds of pupil
# activity, and scales the rest of the data to this baseline.

source("Common.R");

pupilFiles = list.files(path = "../Generated/", pattern = ".pupil.summary$", full.names = TRUE);

for (pupil.file in pupilFiles) {
  print(paste("Reading: ", pupil.file, sep = ""));
  pupil.data = read.table(pupil.file, header = TRUE);

  # Prepare new blank data frame to store output.
  adjustedPupilR = data.frame(
    TimeStamp = numeric(0),
    PupilR = numeric(0),
    Step = character(0)
  );

```

```
  # Calculate baseline pupil diameter as the mean of the first five seconds of recorded activity.
  first5Seconds = subset(pupil.data, TimeStamp < 5, "PupilR");
  baselineR = round(mean(first5Seconds), digits=2);

  for (time in 1:length(pupil.data$PupilR)) {
    # Adjust value as a percent change in radius over the baseline.
    adjustedR = round(((pupil.data$PupilR[time] / baselineR) - 1) * 100, digits=5);

    # Add row to new data frame.
    adjustedRow = data.frame(
        pupil.data$TimeStamp[time],
        adjustedR,
        pupil.data$Step[time]
    );
    colnames(adjustedRow) = colnames(adjustedPupilR);
    adjustedPupilR = rbind(adjustedPupilR, adjustedRow);
  }

  # Write output to a ".pupil.adjusted" file.
  pupil.outputFile = sub(".pupil.summary", ".pupil.adjusted", pupil.file);
  print(paste("Writing: ", pupil.outputFile, sep = ""));
  write.table(adjustedPupilR, file = pupil.outputFile, row.names = FALSE, quote = FALSE);
}
```

7.10.3 PupilSummarizer.R

```
# The raw pupil data as well as smoothened pupil data contains data at 30 Hz, which is far too
# too much to draw a graph from. (The PDF renderer crashes when drawing a graph that contains
# as many data points.) This script summarizes the graph by generating one reading per second,
# which is calculated as the mean of all the readings taken within that second. This data is
# then used to plot all graphs.

source("Common.R");

pupilFiles = list.files(path = "../Generated/", pattern = ".pupil.smooth$", full.names = TRUE);

for (pupil.file in pupilFiles) {
  print(paste("Reading: ", pupil.file));
  pupil.data = read.table(pupil.file, header = TRUE);

  # Create a new data frame to store the results.
  pupilSummary = data.frame(
      TimeStamp = numeric(0),
      PupilR = numeric(0),
      PupilRStdDev = numeric(0),
      Step = character(0)
  );

  # The loop counter must run once for each second.
  maxTimeStamp = ceiling(max(pupil.data$Time));
  for (time in 1:maxTimeStamp) {
```

```
    # Get all pupil measurements during this complete second.
    pupil1SecInterval = subset(pupil.data, TimeStamp > (time - 1) & TimeStamp <= time);
    if (length(pupil1SecInterval$Time) == 0) {
      next;
    }

    # Also calculate standard deviations.
    # SD = 0 if there's only one observation in that second, but R will give us an NA.
    sdPupilR = round(sd(pupil1SecInterval$PupilR), digits = 3);
    if (is.na(sdPupilR)) {
      sdPupilR = 0;
    }

    # Get Step# of the first reading within this second.
    step = min(pupil1SecInterval$Step);

    # Add row to new data frame.
    summaryRow = data.frame(
        time,
        round(mean(pupil1SecInterval$PupilR), digits = 3),
        sdPupilR,
        step);
    colnames(summaryRow) = colnames(pupilSummary);
    pupilSummary = rbind(pupilSummary, summaryRow);
  }

  # Write output to a ".pupil.summary" file.
  pupil.outputFile = sub(".pupil.smooth", ".pupil.summary", pupil.file);
  print(paste("Writing: ", pupil.outputFile, sep = ""));
  write.table(pupilSummary, file = pupil.outputFile, row.names = FALSE, quote = FALSE);
}
```

7.10.4 PupilRawSmoothGraphs.R

```
# This script draws graphs showing an example of raw pupil data (60 s sample)
# and the same data after applying a Savitzky-Golay filter.

library(gplots);
source("Common.R");

# For printing
pdf(width = 8, height = 7, pointsize = 18, fonts = c("MyriadPro"));

# For slides
# par(fg='white', col='white', col.axis='white', col.lab='white',
#     col.main='transparent', col.sub='white', lwd=2, fonts = c("MyriadPro"));

raw.file = "../Samples/RawPupil.pupil";
smooth.file = "../Samples/SmoothPupil.smooth";

raw.data = read.table(raw.file, header = TRUE);
```

```
smooth.data = read.table(smooth.file, header = TRUE);

raw.data.subset = subset(raw.data, TimeStamp >= 60 & TimeStamp < 120);
smooth.data.subset = subset(smooth.data, TimeStamp >= 60 & TimeStamp < 120);

plot(x = raw.data.subset$TimeStamp,
   y = raw.data.subset$PupilR,
   xlab = "Time Elapsed (seconds)",
   ylab = "Pupil Radius (eye image pixels)",
   ylim = c(40, 65),
   type = "l",
   main = "Pupil Data before Smoothing (60 s sample)",
   family = "MyriadPro"
);

plot(x = smooth.data.subset$TimeStamp,
   y = smooth.data.subset$PupilR,
   xlab = "Time Elapsed (seconds)",
   ylab = "Smoothed Pupil Radius",
   ylim = c(40, 65),
   type = "l",
   main = "Pupil Data after Smoothing (60 s sample)",
   family = "MyriadPro"
);
```

7.10.5 TLX.R

```
# This script detects differences in TLX ratings for Levels L0 and L1 for each task
# for the three tasks, Files, Calendar, and Contacts.

library(gplots);
source("Common.R");

pdf(width = 8, height = 6, pointsize = 12, fonts = c("MyriadPro"));

# Look at Data Summaries first
tlx.file = "../PIM Study/NASATLX-Scores.csv";
tlx.data = read.table(tlx.file, header = TRUE, sep = ",", quote = "");

# Put the factors in the order we want, Files, Calendar, then Contacts.
tlx.data$Task = factor(as.character(tlx.data$Task), levels=c("Files", "Calendar", "Contacts"));

measures = c("MD", "PD", "TD", "OP", "EF", "FR", "OverallWorkload");

for (measure in measures) {
  print("------- ANOVA --------");

  anovaFormula = as.formula(paste(measure, " ~ Task", sep=""));
  print(anovaFormula);

  measure.anova = aov(anovaFormula, data = tlx.data);
```

```
  print(summary(measure.anova));
  print(model.tables(measure.anova, "means"), digits = 3);

  # Get standard deviations.
  print("------- Means and SDs --------");

  for (task in c("Files", "Calendar", "Contacts")) {
    for (level in c("L0", "L1")) {
      tlx.data.perLevel = subset(tlx.data, Treatment==level & Task==task, select=measure);
      print(paste("Mean (SD): ", task, measure, level));
      print(paste(
          round(mean(tlx.data.perLevel), digits=3),
          " (",
          round(sd(tlx.data.perLevel), digits=3),
          ")",
          sep = ""
      ));
    }
  }

  # Tukey HSD Post-Hoc
  print("-------Tukey's HSD--------");
  tukeyHsd = TukeyHSD(measure.anova);
  print(tukeyHsd);

  # Draw boxplot; Everything in one graph => Easier comparisons.
  boxplotFormula = as.formula(paste(measure, " ~ Treatment * Task", sep=""));

  plotName = paste(measure, " versus Treatment", sep="");
  par(family = "MyriadPro");
  boxplot(boxplotFormula, data = tlx.data,
      boxwex = 0.5,
      col = lineColors,
      main = plotName,
      xlab = "Treatment Levels",
      ylab = paste(metric.name(measure), " Rating", sep=""),
      ylim = c(0,100),
      family = "MyriadPro",
      axes = FALSE
  );

  axis(1, family = "MyriadPro",
      at = 1:6,
      labels = c("Files L0", "Files L1", "Calendar L0", "Calendar L1", "Contacts L0", "Contacts L1"));
  axis(2, family = "MyriadPro");
  smartlegend(x = "right", y = "top", inset = 0.01,
       c("L0", "L1"),
       fill = lineColors);
}
```

7.10.6 TimePerStep.R

```
# This script performs an ANalysis Of VAriance (ANOVA) of the time taken per each step of each task.
# It also draws graphs showing trends in time taken, superimposed for Level L0 and L1 in different
# colors.

library(gplots);
source("Common.R");

pdf(width = 10, height = 7, pointsize = 12, fonts = c("MyriadPro"));

# Open data summaries.
timing.file = "../Generated/TimePerStep.gen";
timing.data = read.table(timing.file, header = TRUE, sep = "\t", quote = "");

# Define a function that will do the calculations, then call it once for each of the three tasks.
plotTimingForTask = function(task, howManySteps, labels) {
  # Create a new blank data frame to store the mean time and SD for each step.
  timingMeans = data.frame(
      Treatment = character(0),
      Step = character(0),
      MeanTime = numeric(0),
      SD = numeric(0));

  for (loopTreatment in c("L0", "L1")) {
    for (loopStep in 1:howManySteps) {
      timingSec = subset(timing.data, Task==task & Step==loopStep & Treatment==loopTreatment);
      print(timingSec);

      # Append a row for each step, for each level.
      meansRow = data.frame(loopTreatment, loopStep, mean(timingSec$Time), sd(timingSec$Time));
      colnames(meansRow) = colnames(timingMeans);
      timingMeans = rbind(timingMeans, meansRow);
      print(meansRow);
    }
  }

  print(timingMeans);

  # Plot a graph of the timing means for L0, in one color.
  timingMeansL0 = subset(timingMeans, Treatment=="L0");
  plotCI(
      x = timingMeansL0$Step,
      y = timingMeansL0$MeanTime,
      uiw = timingMeansL0$SD,
      lty = "solid",
      lwd = 4,
      pch = 22,
      # xaxt ="n",
      gap = 0,
      col = "red",
```

```
    type = "o",
    xlab = "Step #",
    ylab = "Time Taken (s)",
    main = paste("Time on ", task, " Task", sep=""),
    axes = FALSE,
    family = "MyriadPro"
);

# Plot a graph of the timing means for L1, in a different color, and superimpose it.
timingMeansL1 = subset(timingMeans, Treatment=="L1");
plotCI(
    x = timingMeansL1$Step,
    y = timingMeansL1$MeanTime,
    uiw = timingMeansL1$SD,
    add = TRUE,
    lty = "dashed",
    lwd = 3,
    xaxt ="n",
    col = "green",
    type = "o",
    gap = 0,
    xlab = "",
    ylab = "",
    main = "",
    axes = FALSE,
    family = "MyriadPro"
);

# Draw the axes.
axis(1, at = 1:length(timingMeans$Step), labels = timingMeans$Step, family = "MyriadPro");
axis(2, family = "MyriadPro");
smartlegend(
  x = "right",
  y = "top",
  labels,
  fill = c("red", "green")
  # , family = "MyriadPro"
);

# Perform an ANOVA for each step, see whether there are significant differences in time
# taken for each step between L0 and L1.
for (loopStep in 1:howManySteps) {
  timingSubset = subset(timing.data, Task==task & Step==loopStep);
  timing.anova = aov(Time ~ Treatment, data = timingSubset);

  print(task);
  print(loopStep);
  print(timingSubset);
  print(summary(timing.anova));

  timingSubset.l0 = subset(timingSubset, Treatment=="L0");
```

```
    timingSubset.l1 = subset(timingSubset, Treatment=="L1");

    # Calculate Cohen's d for effect size.
    cohensD = cohens.d(
      mean(timingSubset.l0$Time),
      sd(timingSubset.l0$Time),
      length(timingSubset.l0$Time),
      mean(timingSubset.l1$Time),
      sd(timingSubset.l1$Time),
      length(timingSubset.l1$Time)
    );
    print(paste("Cohen's d (effect size) = ", cohensD));
  }

  boxplot(Time ~ Treatment * Step,
    data = subset(timing.data, Task==task),
    boxwex = 0.5,
    ylim = c(0,500),
    col = colors
  );
}

# Now call the function we just created for each of the three tasks.
plotTimingForTask("Files", 10, c("Without Sync Support", "With Sync Support"));
plotTimingForTask("Calendar", 16, c("Paper Calendar", "Online Calendar"));
plotTimingForTask("Contacts", 7, c("No Sync Support", "Sync Support"));
```

7.10.7 PupilANOVAPerStep.R

```
# This script performs two analyses on Pupil Radius on a step-wise basis.
# 1. ANOVA to detect differences between corresponding steps of the task at L0 and L1;
# 2. ANOVA to detect differences among steps within the same task execution (either only L0 or only L1)

library(gplots);
source("Common.R");

pdf(width = 10, height = 8, pointsize = 12, fonts = c("MyriadPro"));

stepwise.workload.file = "../Generated/Stepwise.workload";
stepwise.workload.data = read.table(stepwise.workload.file, header = TRUE);

tasks = c("Files", "Calendar", "Contacts");

for (task in tasks) {
  # Now prepare an auxiliary table that has "Step" & "mean(PupilR)" as its two columns.
  stepwisePupilMeans = data.frame(
      Task = character(0),
      Treatment = character(0),
      Step = character(0),
      MeanPupilR = numeric(0),
      SDPupilR = numeric(0));
```

```

# Do an ANOVA for each step, comparing pupil radius for L0 and L1
for (step in 1:(stepsForTask(task)-1)) {
  title = paste(task, "Task, Step", step);
  print(title);

  stepwiseWorkloadPerTaskPerStep = subset(stepwise.workload.data, Task==task & Step==step);
  anovaFormula = PupilR ~ Treatment;

  stepwise.workload.anova = aov(anovaFormula, data = stepwiseWorkloadPerTaskPerStep);
  print(summary(stepwise.workload.anova));
  print(model.tables(stepwise.workload.anova, "means"), digits = 3);

  # Now prepare an auxiliary table that has "Step" & "mean(PupilR)" as its two columns.
  for (level in c("L0", "L1")) {
    workloadPerTaskPerLevelPerStep = subset(stepwiseWorkloadPerTaskPerStep, Treatment==level);

    stepwisePupilMeansRow = data.frame(
        task,
        level,
        step,
        round(mean(workloadPerTaskPerLevelPerStep$PupilR), digits = 5),
        round(sd(workloadPerTaskPerLevelPerStep$PupilR), digits = 5));
    colnames(stepwisePupilMeansRow) = colnames(stepwisePupilMeans);
    stepwisePupilMeans = rbind(stepwisePupilMeans, stepwisePupilMeansRow);
  }
}

# Now plot a chart for each step together in one graph.
stepwisePupilMeansL0 = subset(stepwisePupilMeans, Treatment=="L0");
stepwisePupilMeansL1 = subset(stepwisePupilMeans, Treatment=="L1");

plotCI(
    x = stepwisePupilMeansL0$Step,
    y = stepwisePupilMeansL0$MeanPupilR,
    uiw = stepwisePupilMeansL0$SDPupilR,
    lty = "dashed",
    lwd = 3,
    col = "red",
    type = "o",
    family = "MyriadPro",
    main = paste(task, "Task"),
    xlab = "Step within Task",
    ylab = "Adjusted Pupil Radius",
    axes = FALSE);

plotCI(
    x = stepwisePupilMeansL1$Step,
    y = stepwisePupilMeansL1$MeanPupilR,
    uiw = stepwisePupilMeansL0$SDPupilR,
    lty = "dashed",
```

```
      lwd = 3,
      col = "green",
      xaxt ="n",
      type = "o",
      add = TRUE,
      main = "",
      family = "MyriadPro",
      axes = FALSE);

  axis(1, at = 1:length(stepwisePupilMeansL1$Step), labels = stepwisePupilMeansL1$Step,
      family = "MyriadPro");
  axis(2, family = "MyriadPro");
  smartlegend(
      x = "right",
      y = "top",
      c("L0", "L1"),
      fill = c("red", "green"));

  # Do a second analysis: ANOVA among all steps in one level to detect differences in workload
  # among levels.
  for (level in c("L0", "L1")) {
    print(paste(task, level));

    stepwise.workload.data$Step = factor(as.character(stepwise.workload.data$Step));
    workloadAtLevel = subset(stepwise.workload.data, Task==task & Treatment==level);

    # print(workloadAtLevel);
    print("ANOVA among all steps in one level to detect differences in workload among levels.");
    workloadAtLevel.anova = aov(PupilR ~ Step, data = workloadAtLevel);
    print(summary(workloadAtLevel.anova));
    print(model.tables(workloadAtLevel.anova, "means"), digits = 3);
    print(TukeyHSD(workloadAtLevel.anova));
  }
}
```

7.10.8 PupilGraphs.R

```
library(gplots);
source("Common.R");

pdf(width = 11, height = 8.5, pointsize = 12, fonts = c("MyriadPro"));

pupilFiles = list.files(path = "../Generated/", pattern = ".*Files\\.pupil.adjusted$", full.names = TRUE);
for (pupil.file in pupilFiles) {
  print(paste("Reading ", pupil.file));

  participant = sub("^.*(P[[:digit:]]{1,2}).*$", '\\1', pupil.file);
  level = sub("^.*(L[[:digit:]]{1}).*$", '\\1', pupil.file);
  task = sub("^.* ([[:alpha:]]*).pupil.adjusted$", '\\1', pupil.file);

  timing.file = sub(".pupil.adjusted", ".timing", pupil.file);
```

```
  pdf.file = sub(".gen", ".pdf", sub("/Generated/", "/Graphs/", pupil.file));
  print(paste("Writing ", pdf.file));

  pupil.data = read.table(pupil.file, header = TRUE);
  timing.data = read.table(timing.file, header = TRUE);

  pupilDataMinusStepZero = subset(pupil.data, Step!=0);

  # Plot the eye-tracker data.
  plot(x = pupilDataMinusStepZero$TimeStamp,
      y = pupilDataMinusStepZero$PupilR,
      xlab = "Time Elapsed (seconds)",
      ylab = "Pupil Radius (eye image pixels)",
      ylim = c(-30, 30),
      type = "l",
      main = paste(task, " Task", ", Participant ", participant, ", Level ", level,   sep=""),
      family = "MyriadPro");

  # Draw a vertical line at each step.
  # Draw it before the pupil data, so its z-index is lower.
  abline(v = timing.data$Time, col = rgb(0.4,0.8,1), lwd = 2, family = "MyriadPro");

  # Draw a horizontal line at adjusted pupil size = 0.
  abline(h = 0, col = "gray", lwd = 1, family = "MyriadPro");

  # Write the Step # as superimposed text.
  for (i in 1:length(timing.data$Time) - 1) {
    label = paste("S", timing.data$Step[i], sep="");
    text(x = as.numeric(timing.data$Time[i]) - 5, y = -1, labels = label, pos = 4,
        col = lineColors[1], family = "MyriadPro");
  }
}
```

7.11 Creative Commons Legal Code

THE WORK (AS DEFINED BELOW) IS PROVIDED UNDER THE TERMS OF THIS CREATIVE COMMONS PUBLIC LICENSE ("CCPL" OR "LICENSE"). THE WORK IS PROTECTED BY COPYRIGHT AND/OR OTHER APPLICABLE LAW. ANY USE OF THE WORK OTHER THAN AS AUTHORIZED UNDER THIS LICENSE OR COPYRIGHT LAW IS PROHIBITED.

BY EXERCISING ANY RIGHTS TO THE WORK PROVIDED HERE, YOU ACCEPT AND AGREE TO BE BOUND BY THE TERMS OF THIS LICENSE. TO THE EXTENT THIS LICENSE MAY BE CONSIDERED TO BE A CONTRACT, THE LICENSOR GRANTS YOU THE RIGHTS CONTAINED HERE IN CONSIDERATION OF YOUR ACCEPTANCE OF SUCH TERMS AND CONDITIONS.

Definitions

1. **"Collective Work"** means a work, such as a periodical issue, anthology or encyclopedia, in which the Work in its entirety in unmodified form, along with one or more other contributions, constituting separate and independent works in themselves, are assembled into a collective whole. A work that constitutes a Collective Work will not be considered a Derivative Work (as defined below) for the purposes of this License.

2. **"Derivative Work"** means a work based upon the Work or upon the Work and other preexisting works, such as a translation, musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which the Work may be recast, transformed, or adapted, except that a work that constitutes a Collective Work will not be considered a Derivative Work for the purpose of this License. For the avoidance of doubt, where the Work is a musical composition or sound recording, the synchronization of the Work in timed-relation with a moving image ("synching") will be considered a Derivative Work for the purpose of this License.

3. **"Licensor"** means the individual, individuals, entity or entities that offer(s) the Work under the terms of this License.

4. **"Original Author"** means the individual, individuals, entity or entities who created the Work.

5. **"Work"** means the copyrightable work of authorship offered under the terms of this License.

6. **"You"** means an individual or entity exercising rights under this License who has not previously violated the terms of this License with respect to the Work, or who has received express permission from the Licensor to exercise rights under this License despite a previous violation.

7. **"License Elements"** means the following high-level license attributes as selected by Licensor and indicated in the title of this License: Attribution, Noncommercial, ShareAlike.

Fair Use Rights.

Nothing in this license is intended to reduce, limit, or restrict any rights arising from fair use, first sale or other limitations on the exclusive rights of the copyright owner under copyright law or other applicable laws.

License Grant.

Subject to the terms and conditions of this License, Licensor hereby grants You a worldwide, royalty-free, non-exclusive, perpetual (for the duration of the applicable copyright) license to exercise the rights in the Work as stated below:

1. to reproduce the Work, to incorporate the Work into one or more Collective Works, and to reproduce the Work as incorporated in the Collective Works;
2. to create and reproduce Derivative Works provided that any such Derivative Work, including any translation in any medium, takes reasonable steps to clearly label, demarcate or otherwise identify that changes were made to the original Work. For example, a translation could be marked "The original work was translated from English to Spanish," or a modification could indicate "The original work has been modified.";
3. to distribute copies or phonorecords of, display publicly, perform publicly, and perform publicly by means of a digital audio transmission the Work including as incorporated in Collective Works;
4. to distribute copies or phonorecords of, display publicly, perform publicly, and perform publicly by means of a digital audio transmission Derivative Works;

The above rights may be exercised in all media and formats whether now known or hereafter devised. The above rights include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. All rights not expressly granted by Licensor are hereby reserved, including but not limited to the rights set forth in Sections 4(e) and 4(f).

Restrictions.

The license granted in Section 3 above is expressly made subject to and limited by the following restrictions:

1. You may distribute, publicly display, publicly perform, or publicly digitally perform the Work only under the terms of this License, and You must include a copy of, or the Uniform Resource Identifier for, this License with every copy or phonorecord of the Work You distribute, publicly display, publicly perform, or publicly digitally perform. You may not offer or impose any terms on the Work that restrict the terms of this License or the ability of a recipient of the Work to exercise the rights granted to that recipient under the terms of the License. You may not sublicense the Work. You must keep intact all notices that refer to this License and to the disclaimer of warranties. When You distribute, publicly display, publicly perform, or publicly digitally perform the Work, You may not impose any technological measures on the Work that restrict the ability of a recipient of the Work from You to exercise the rights granted to that recipient under the terms of the License. This Section 4(a) applies to the Work as incorporated in a Collective Work, but this does not require the Collective Work apart from the Work itself to be made subject to the terms of this License. If You create a Collective Work, upon notice from any Licensor You must, to the extent practicable, remove from the Collective Work any credit as required by Section 4(d), as requested. If You create a Derivative Work, upon notice from any Licensor You must, to the extent practicable, remove from the Derivative Work any credit as required by Section 4(d), as requested.

2. You may distribute, publicly display, publicly perform, or publicly digitally perform a Derivative Work only under: (i) the terms of this License; (ii) a later version of this License with the same License Elements as this License; or, (iii) either the unported Creative Commons license or a Creative Commons license for another jurisdiction (either this or a later license version) that contains the same License Elements as this License (e.g. Attribution-NonCommercial-ShareAlike 3.0 (Unported)) ("the Applicable License"). You must include a copy of, or the Uniform Resource Identifier for, the Applicable License with every copy or phonorecord of each Derivative Work You distribute, publicly display, publicly perform, or publicly digitally perform. You may not offer or impose any terms on the Derivative Works that restrict the terms of the Applicable License or the ability of a recipient of the Work to exercise the rights granted to that recipient under the terms of the Applicable License. You must keep intact all notices that refer to the Applicable License and to the disclaimer of warranties. When You distribute, publicly display, publicly perform, or publicly digitally perform the Derivative Work, You may not impose any technological measures on the Derivative Work that restrict the ability of a recipient of the Derivative Work from You to exercise the rights granted to that recipient under the terms of the Applicable License. This Section 4(b) applies to the Derivative Work as incorporated in a Collective Work, but this does not require the Collective Work apart from the Derivative Work itself to be made subject to the terms of the Applicable License.

3. You may not exercise any of the rights granted to You in Section 3 above in any manner that is primarily intended for or directed toward commercial advantage or private monetary

compensation. The exchange of the Work for other copyrighted works by means of digital file-sharing or otherwise shall not be considered to be intended for or directed toward commercial advantage or private monetary compensation, provided there is no payment of any monetary compensation in connection with the exchange of copyrighted works.

4. If You distribute, publicly display, publicly perform, or publicly digitally perform the Work (as defined in Section 1 above) or any Derivative Works (as defined in Section 1 above) or Collective Works (as defined in Section 1 above), You must, unless a request has been made pursuant to Section 4(a), keep intact all copyright notices for the Work and provide, reasonable to the medium or means You are utilizing: (i) the name of the Original Author (or pseudonym, if applicable) if supplied, and/or (ii) if the Original Author and/or Licensor designate another party or parties (e.g. a sponsor institute, publishing entity, journal) for attribution ("Attribution Parties") in Licensor's copyright notice, terms of service or by other reasonable means, the name of such party or parties; the title of the Work if supplied; to the extent reasonably practicable, the Uniform Resource Identifier, if any, that Licensor specifies to be associated with the Work, unless such URI does not refer to the copyright notice or licensing information for the Work; and, consistent with Section 3(b) in the case of a Derivative Work, a credit identifying the use of the Work in the Derivative Work (e.g., "French translation of the Work by Original Author," or "Screenplay based on original Work by Original Author"). The credit required by this Section 4(d) may be implemented in any reasonable manner; provided, however, that in the case of a Derivative Work or Collective Work, at a minimum such credit will appear, if a credit for all contributing authors of the Derivative Work or Collective Work appears, then as part of these credits and in a manner at least as prominent as the credits for the other contributing authors. For the avoidance of doubt, You may only use the credit required by this Section for the purpose of attribution in the manner set out above and, by exercising Your rights under this License, You may not implicitly or explicitly assert or imply any connection with, sponsorship or endorsement by the Original Author, Licensor and/or Attribution Parties, as appropriate, of You or Your use of the Work, without the separate, express prior written permission of the Original Author, Licensor and/or Attribution Parties.

5. For the avoidance of doubt, where the Work is a musical composition:

   (a) **Performance Royalties Under Blanket Licenses.** Licensor reserves the exclusive right to collect whether individually or, in the event that Licensor is a member of a performance rights society (e.g. ASCAP, BMI, SESAC), via that society, royalties for the public performance or public digital performance (e.g. webcast) of the Work if that performance is primarily intended for or directed toward commercial advantage or private monetary compensation.

   (b) **Mechanical Rights and Statutory Royalties.** Licensor reserves the exclusive right to collect, whether individually or via a music rights agency or designated agent (e.g. Harry Fox Agency), royalties for any phonorecord You create from the Work ("cover version") and distribute, subject to the compulsory license created by 17 USC Section 115 of the US Copyright Act (or the equivalent in other jurisdictions), if Your distribution of such cover version is primarily intended for or directed toward commercial advantage or private monetary compensation.

6. **Webcasting Rights and Statutory Royalties.** For the avoidance of doubt, where the Work is a sound recording, Licensor reserves the exclusive right to collect, whether individually or via a performance-rights society (e.g. SoundExchange), royalties for the public digital performance (e.g. webcast) of the Work, subject to the compulsory license created by 17 USC Section 114 of the US Copyright Act (or the equivalent in other jurisdictions), if Your public digital performance is primarily intended for or directed toward commercial advantage or private monetary compensation.

Representations, Warranties and Disclaimer

UNLESS OTHERWISE MUTUALLY AGREED TO BY THE PARTIES IN WRITING, LICENSOR OFFERS THE WORK AS-IS AND ONLY TO THE EXTENT OF ANY RIGHTS HELD IN THE LICENSED WORK BY THE LICENSOR. THE LICENSOR MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND CONCERNING THE WORK, EXPRESS, IMPLIED, STATUTORY OR OTHERWISE, INCLUDING, WITHOUT LIMITATION, WARRANTIES OF TITLE, MARKETABILITY, MERCHANTIBILITY, FITNESS FOR A PARTICULAR PURPOSE, NONINFRINGEMENT, OR THE ABSENCE OF LATENT OR OTHER DEFECTS, ACCURACY, OR THE PRESENCE OF ABSENCE OF ERRORS, WHETHER OR NOT DISCOVERABLE. SOME JURISDICTIONS DO NOT ALLOW THE EXCLUSION OF IMPLIED WARRANTIES, SO SUCH EXCLUSION MAY NOT APPLY TO YOU.

Limitation on Liability.

EXCEPT TO THE EXTENT REQUIRED BY APPLICABLE LAW, IN NO EVENT WILL LICENSOR BE LIABLE TO YOU ON ANY LEGAL THEORY FOR ANY SPECIAL, INCIDENTAL, CONSEQUENTIAL, PUNITIVE OR EXEMPLARY DAMAGES ARISING OUT OF THIS LICENSE OR THE USE OF THE WORK, EVEN IF LICENSOR HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

Termination

1. This License and the rights granted hereunder will terminate automatically upon any breach by You of the terms of this License. Individuals or entities who have received Derivative Works (as defined in Section 1 above) or Collective Works (as defined in Section 1 above) from You under this License, however, will not have their licenses terminated provided such individuals or entities remain in full compliance with those licenses. Sections 1, 2, 5, 6, 7, and 8 will survive any termination of this License.

2. Subject to the above terms and conditions, the license granted here is perpetual (for the duration of the applicable copyright in the Work). Notwithstanding the above, Licensor reserves the right to release the Work under different license terms or to stop distributing the Work at any time; provided, however that any such election will not serve to withdraw this License (or any other license that has been, or is required to be, granted under the terms of this License), and this License will continue in full force and effect unless terminated as stated above.

Miscellaneous

1. Each time You distribute or publicly digitally perform the Work (as defined in Section 1 above) or a Collective Work (as defined in Section 1 above), the Licensor offers to the recipient a license to the Work on the same terms and conditions as the license granted to You under this License.

2. Each time You distribute or publicly digitally perform a Derivative Work, Licensor offers to the recipient a license to the original Work on the same terms and conditions as the license granted to You under this License.

3. If any provision of this License is invalid or unenforceable under applicable law, it shall not affect the validity or enforceability of the remainder of the terms of this License, and without further action by the parties to this agreement, such provision shall be reformed to the minimum extent necessary to make such provision valid and enforceable.

4. No term or provision of this License shall be deemed waived and no breach consented to unless such waiver or consent shall be in writing and signed by the party to be charged with such waiver or consent.

5. This License constitutes the entire agreement between the parties with respect to the Work licensed here. There are no understandings, agreements or representations with respect to the Work not specified here. Licensor shall not be bound by any additional provisions that may appear in any communication from You. This License may not be modified without the mutual written agreement of the Licensor and You.

Bibliography

[Abowd and Mynatt, 2000] Abowd, G. D. and Mynatt, E. D. (2000). Charting past, present, and future research in ubiquitous computing. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 7(1):29–58.

[Abrams et al., 1998] Abrams, D., Baecker, R., and Chignell, M. (1998). Information archiving with bookmarks: Personal web space construction and organization. In *CHI '98: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 41–48, New York, NY, USA. ACM Press/Addison-Wesley Publishing Co.

[Abrams et al., 1999] Abrams, M., Phanouriou, C., Batongbacal, A., Williams, S., and Shuster, J. (1999). UIML: An appliance-independent XML User Interface Language. In *Proceedings of the 8th WWW conference*.

[Adams, 2002] Adams, D. (2002). *The Salmon of Doubt: Hitchhiking the Galaxy One Last Time*. Harmony.

[Adar et al., 1999] Adar, E., Karger, D., and Stein, L. A. (1999). Haystack: Per-user information environments. In *CIKM '99: Proceedings of the 8th International Conference on Information and Knowledge Management*, pages 413–422, New York, NY, USA. ACM Press.

[Ali et al., 2005] Ali, M. F., Pérez-Quiñones, M., and Abrams, M. (2005). *Building Multi-Platform User Interfaces with UIML*, pages 93–118. Multiple User Interfaces. John Wiley & Sons, Ltd.

[Apple, 2004] Apple (2004). Mac OS X Spotlight. http://www.apple.com/macosx/features/spotlight/.

[Applied Science Laboratories, 2007] Applied Science Laboratories (2007). *MobileEye Operation Manual*. Applied Science Group Company, Bedford, MA, 1.32 edition.

[Backs and Walrath, 1992] Backs, R. W. and Walrath, L. C. (1992). Eye movement and pupillary response indices of mental workload during visual search of symbolic displays. *Applied Ergonomics*, 23(4):243–254.

[Bagozzi, 1992] Bagozzi, R. P. (1992). The self-regulation of attitudes, intentions, and behavior. *Social Psychology Quarterly*, 55(2):178–204.

[Baguley, 2004] Baguley, T. (2004). Understanding statistical power in the context of applied research. *Applied Ergonomics*, 35(2):73–80.

[Bailey and Iqbal, 2008] Bailey, B. P. and Iqbal, S. T. (2008). Understanding changes in mental workload during execution of goal-directed tasks and its application for interruption management. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 14(4):1–28.

[Ballas et al., 1992a] Ballas, J., Heitmeyer, C., and Pérez-Quiñones, M. (1992a). Evaluating two aspects of direct manipulation in advanced cockpits. In *CHI '92: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 127–134, New York, NY, USA. ACM.

[Ballas et al., 1992b] Ballas, J. A., Heitmeyer, C. L., and Pérez-Quiñones, M. (1992b). Direct manipulation and intermittent automation in advanced cockpits. Technical Report NRL/FR/5534-92-9375, Naval Research Laboratory, Arlington, VA.

[Bälter and Sidner, 2002] Bälter, O. and Sidner, C. (2002). Bifrost inbox organizer: Giving users control over the inbox. In *NordiCHI '02: Proceedings of the Second Nordic Conference on Human-Computer Interaction*, pages 111–118, New York, NY, USA. ACM Press.

[Bandura, 2000] Bandura, A. (2000). Exercise of human agency through collective efficacy. *Current Directions in Psychological Science*, 9(3):75–78.

[Barreau, 1995] Barreau, D. (1995). Context as a factor in personal information management systems. *Journal of the American Society for Information Science*, 46(5):327–339.

[Barreau and Nardi, 1995] Barreau, D. and Nardi, B. A. (1995). Finding and reminding: file organization from the desktop. *SIGCHI Bulletin*, 27(3):39–43.

[Battiste and Bortolussi, 1988] Battiste, V. and Bortolussi, M. (1988). Transport pilot workload: A comparison of two subjective techniques. In *Proceedings of the Human Factors Society Thirty-Second Annual Meeting*, pages 150–154, Santa Monica, CA. Human Factors Society.

[Beatty, 1982] Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, 91(2):276–92.

[Bellotti et al., 2003] Bellotti, V., Ducheneaut, N., Howard, M., and Smith, I. (2003). Taking email to task: the design and evaluation of a task management centered email tool. In *CHI '03: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 345–352, New York, NY, USA. ACM Press.

[Bellotti et al., 2002] Bellotti, V., Ducheneaut, N., Howard, M., Smith, I., and Neuwirth, C. (2002). Innovation in extremis: Evolving an application for the critical work of email and information management. In *DIS '02: Proceedings of the Conference on Designing Interactive Systems*, pages 181–192, New York, NY, USA. ACM Press.

[Bellotti and Smith, 2000] Bellotti, V. and Smith, I. (2000). Informing the design of an information management system with iterative fieldwork. In *DIS '00: Proceedings of the Conference on Designing Interactive Systems*, pages 227–237, New York, NY, USA. ACM Press.

[Bellotti and Thornton, 2006] Bellotti, V. and Thornton, J. (2006). Managing activities with TV-ACTA. In *Proceedings of the 2nd Invitational Workshop on Personal Information Management at SIGIR 2006*.

[Bergman et al., 2003] Bergman, O., Beyth-Marom, R., and Nachmias, R. (2003). The user-subjective approach to personal information management systems. *Journal of the American Society for Information Science and Technology*, 54(9):872–878.

[Bergman et al., 2006] Bergman, O., Beyth-Marom, R., and Nachmias, R. (2006). The project fragmentation problem in personal information management. In *CHI '06: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 271–274, New York, NY, USA. ACM Press.

[Bergman et al., 2008] Bergman, O., Beyth-Marom, R., and Nachmias, R. (2008). The user-subjective approach to personal information management systems design: Evidence and implementations. *Journal of the American Society for Information Science and Technology*, 59(2):235–246.

[Bernstein et al., 2008] Bernstein, M., Kleek, M. V., mc schraefel, and Karger, D. (2008). Evolution and evaluation of an information scrap manager. In *3rd Invitational Workshop on Personal Information Managementat CHI 2008: The Disappearing Desktop (PIM 2008)*.

[Bertram et al., 1992] Bertram, D. A., Opila, D. A., Brown, J. L., Gallagher, S. J., Schifeling, R. W., Snow, I. S., and Hershey, C. O. (1992). Measuring physician mental workload: Reliability and validity assessment of a brief instrument. *Medical Care*, 30(2):95–104.

[Boardman and Sasse, 2004] Boardman, R. and Sasse, M. A. (2004). "Stuff goes into the computer and doesn't come out": A Cross-tool study of Personal Information Management. In *CHI '04: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 583–590, New York, NY, USA. ACM Press.

[Boardman et al., 2003] Boardman, R., Spence, R., and Sasse, M. A. (2003). Too many hierarchies?: The daily struggle for control of the workspace. In *Proc. HCI International 2003*.

[Bødker, 1989] Bødker, S. (1989). A human activity approach to user interfaces. *Human-Computer Interaction*, 4(3):171–195.

[Borchers, 2004] Borchers, H. W. (2004). Savitzky-Golay smoothing – an R implementation. https://stat.ethz.ch/pipermail/r-help/2004-February/045568.html.

[Bush, 1945] Bush, V. (1945). As we may think. *The Atlantic Monthly*.

[Butcher, 1995] Butcher, H. (1995). Information overload in management and business. In *IEE Colloquium Digest*, volume 95, pages 1–2, London.

[Capra, 2006] Capra, R. G. (2006). *An Investigation of Finding and Refinding Information on the Web*. PhD thesis, Virginia Tech, Blacksburg, VA.

[Capra and Pérez-Quiñones, 2004] Capra, R. G. and Pérez-Quiñones, M. (2004). Mobile refinding of web information using a voice interface. *Computing Research Repository (CoRR)*, cs.HC/0402001.

[Capra et al., 2001] Capra, R. G., Pérez-Quiñones, M., and Ramakrishnan, N. (2001). WebContext: Remote access to shared context. In *PUI '01: Proceedings of the Perceptual User Interfaces Workshop (PUI 2001)*, pages 1–9, New York, NY, USA. ACM.

[Champely et al., 2007] Champely, S., Ekstrom, C., and Dalgaard, P. (2007). *Basic power calculations using R: pwr*.

[Chau et al., 2008] Chau, D. H., Myers, B., and Faulring, A. (2008). Feldspar: A system for finding information by association. In *3rd Invitational Workshop on Personal Information Managementat CHI 2008: The Disappearing Desktop (PIM 2008)*.

[Chhatpar and Pérez-Quiñones, 2003] Chhatpar, C. and Pérez-Quiñones, M. (2003). Dialogue mobility across devices. In *ACM Southeast Conference (ACMSE)*, Savannah, Georgia.

[Chirita et al., 2006] Chirita, P., Gaugaz, J., Costache, S., and Nejdl, W. (2006). Context detection on the desktop combining multiple evidences. In *Proceedings of the 2nd Invitational Workshop on Personal Information Management at SIGIR 2006.*

[Chu et al., 2004] Chu, H.-h., Song, H., Wong, C., Kurakake, S., and Katagiri, M. (2004). ROAM, a seamless application framework. *Journal of Systems and Software*, 69(3):209–226.

[Cohen, 1988] Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences (2nd Edition)*. Lawrence Erlbaum.

[Collins and Kay, 2008] Collins, A. and Kay, J. (2008). Collaborative personal information management with shared, interactive tabletops. In *3rd Invitational Workshop on Personal Information Managementat CHI 2008: The Disappearing Desktop (PIM 2008)*.

[Cooper and Harper, 1969] Cooper, G. E. and Harper, R. P. (1969). The use of pilot rating in the evaluation of aircraft handling qualities. Technical Report NASA TN D-5153, National Aeronautics and Space Administration.

[Czerwinski et al., 2006] Czerwinski, M., Gage, D. W., Gemmell, J., Marshall, C. C., Pérez-Quiñones, M., Skeels, M. M., and Catarci, T. (2006). Digital memories in an era of ubiquitous computing and abundant storage. *Communications of the Association for Computing Machinery (CACM)*, 49(1):44–50.

[da Silva, 2005] da Silva, J. S. (2005). Challenges and opportunities in ICT: a European perspective. In *MDM '05: Proceedings of the 6th International Conference on Mobile data management*, pages 327–327, New York, NY, USA. ACM.

[Davis, 1989] Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3):319–340.

[de Waard, 1996] de Waard, D. (1996). *The Measurement of Drivers' Mental Workload*. PhD thesis, University of Groningen, Haren, The Netherlands.

[Dearman and Pierce, 2008] Dearman, D. and Pierce, J. S. (2008). "It's on my other computer!": Computing with multiple devices. In *CHI 2008: Proceedings of the ACM Conference on Human Factors in Computing Systems*, page 767.

[Denis and Karsenty, 2004] Denis, C. and Karsenty, L. (2004). Inter-usability of multi-device systems - a conceptual framework. In Seffah, A. and Javahery, H., editors, *Multiple User Interfaces: Cross-Platform Applications and Context-Aware Interfaces*, pages 373–384. John Wiley and Sons.

[Denning, 2006] Denning, P. J. (2006). Infoglut. *Communications of the Association for Computing Machinery (CACM)*, 49(7):15–19.

[Dillon, 2002a] Dillon, A. (2002a). Beyond usability: process, outcome and affect in human-computer interactions. *Canadian Journal of Library and Information Science*, 26(4):57–69.

[Dillon, 2002b] Dillon, A. (2002b). *HCI and the Digital Library*, pages 457–474. HCI and the Millennium. ACM Press/Addison Wesley, New York.

[Dourish, 2001] Dourish, P. (2001). *Where The Action Is: The Foundations of Embodied Interaction*. MIT Press.

[Drucker, 2006] Drucker, P. F. (2006). *The Effective Executive: The Definitive Guide to Getting the Right Things Done*. Collins Business.

[Ducheneaut and Bellotti, 2001] Ducheneaut, N. and Bellotti, V. (2001). E-mail as habitat: An exploration of embedded personal information management. *Interactions*, 8(5):30–38.

[Dumais et al., 2003] Dumais, S., Cutrell, E., Cadiz, J., Jancke, G., Sarin, R., and Robbins, D. C. (2003). Stuff I've Seen: A system for personal information retrieval and re-use. In *SIGIR '03: Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 72–79, New York, NY, USA. ACM Press.

[Dumais and Gates, 2003] Dumais, S. and Gates, B. (2003). Keynote at COMDEX 2003. Public Address.

[Edmunds and Morris, 2000] Edmunds, A. and Morris, A. (2000). The problem of information overload in business organisations: A review of the literature. *International Journal of Information Management*, 20(1):17–28.

[Eggemeier, 1988] Eggemeier, F. T. (1988). Properties of workload assessment techniques. *Human mental workload*, 52:41–62.

[Eggemeier and Wilson, 1991] Eggemeier, F. T. and Wilson, G. F. (1991). *Performance-based and subjective assessment of workload in multi-task environments*, chapter 10. Multiple-Task Performance. Taylor and Francis.

[Eggemeier et al., 1991] Eggemeier, F. T., Wilson, G. F., Kramer, A. F., and Damos, D. L. (1991). *Workload assessment in multi-task environments*, chapter 9. Multiple-Task Performance. Taylor and Francis.

[Einsenstein et al., 2001] Einsenstein, J., Vanderdonckt, J., and Puerta, A. (2001). Applying model-based techniques to the development of UIs for mobile computers. In *Proceedings IUI'01: International Conference on Intelligent User Interfaces*, pages 69–76. ACM Press.

[Eliot, 1934] Eliot, T. S. (1934). *The Rock*.

[Elsweiler et al., 2006] Elsweiler, D., Ruthven, I., , and Ma, L. (2006). Role of memory in PIM. In *Proceedings of the 2nd Invitational Workshop on Personal Information Management at SIGIR 2006*.

[Endicott et al., 1993] Endicott, J., Nee, J., Harrison, W., and Blumenthal, R. (1993). Quality of life enjoyment and satisfaction questionnaire: A new measure. *Psychopharmacology Bulletin*, 29(2):321.

[Farhoomand and Drury, 2002] Farhoomand, A. F. and Drury, D. H. (2002). Managerial information overload. *Communications of the Association for Computing Machinery (CACM)*, 45(10):127–131.

[Fidel and Pejtersen, 2004] Fidel, R. and Pejtersen, A. (2004). From information behaviour research to the design of information systems: the Cognitive Work Analysis framework. *Information Research*, 10(1):10–1.

[Florins and Vanderdonckt, 2004] Florins, M. and Vanderdonckt, J. (2004). Graceful degradation of user interfaces as a design method for multiplatform systems. In *IUI '04: Proceedings of the 9th International Conference on Intelligent User Interfaces*, pages 140–147, New York, NY, USA. ACM Press.

[Fox et al., 1993] Fox, E. A., Hix, D., Nowell, L. T., Brueni, D. J., Wake, W. C., Heath, L. S., and Rao, D. (1993). Users, user interfaces, and objects: Envision, a digital library. *Journal of the American Society for Information Science*, 44(8):480–491.

[Gemmell et al., 2002] Gemmell, J., Bell, G., Lueder, R., Drucker, S., and Wong, C. (2002). MyLifeBits: fulfilling the Memex vision. In *MULTIMEDIA '02: Proceedings of the Tenth ACM International Conference on Multimedia*, pages 235–238, New York, NY, USA. ACM.

[Google, Inc., 2004] Google, Inc. (2004). Google Desktop. http://desktop.google.com/.

[Granholm et al., 1996] Granholm, E., Asarnow, R. F., Sarkin, A. J., and Dykes, K. L. (1996). Pupillary responses index cognitive resource limitations. *Psychophysiology*, 33(4):457–461.

[Gwizdka, 2000] Gwizdka, J. (2000). Timely reminders: A case study of temporal guidance in PIM and email tools usage. In *CHI '00: Extended Abstracts on Human Factors in Computing Systems*, pages 163–164, New York, NY, USA. ACM Press.

[Gwizdka, 2002] Gwizdka, J. (2002). Reinventing the inbox: Supporting the management of pending tasks in email. In *CHI '02: Extended Abstracts on Human Factors in Computing Systems*, pages 550–551, New York, NY, USA. ACM Press.

[Gwizdka, 2004] Gwizdka, J. (2004). Email task management styles: The cleaners and the keepers. In *CHI '04: Extended Abstracts on Human Factors in Computing Systems*, pages 1235–1238, New York, NY, USA. ACM Press.

[Gwizdka, 2006] Gwizdka, J. (2006). Finding to keep and organize: Personal information collections as context. In *Proceedings of the 2nd Invitational Workshop on Personal Information Management at SIGIR 2006.*

[Harrison and Dourish, 1996] Harrison, S. and Dourish, P. (1996). Re-place-ing space: The roles of place and space in collaborative systems. In *CSCW '96: Proceedings of the 1996 ACM conference on Computer Supported Cooperative Work*, pages 67–76, New York, NY, USA. ACM Press.

[Harrison et al., 2007] Harrison, S., Tatar, D., and Sengers, P. (2007). The three paradigms of HCI. In *Proceedings of alt.chi 2007, ACM Conference on Human Factors in Computing Systems*.

[Hart and Staveland, 1988] Hart, S. G. and Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human Mental Workload*, 1:139–183.

[Hendy et al., 1993] Hendy, K. C., Hamilton, K. M., and Landry, L. N. (December 1993). Measuring subjective workload: When is one scale better than many? *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 35:579–601(23).

[Hess, 1975] Hess, E. H. (1975). *The Tell-Tale Eye: How Your Eyes Reveal Hidden Thoughts and Emotions*. Van Nostrand Reinhold, Cincinnati, OH.

[Hess and Polt, 1964] Hess, E. H. and Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science*, 143(3611):1190–1192.

[Hill et al., 1989] Hill, S., Byers, J., Zaklad, A., and Christ, R. (1989). Subjective workload assessment during 48 continuous hours of LOS-F-H operations. In *Proceedings of the Human Factors Society Thirty-Third Annual Meeting*, pages 1129–1133, Santa Monica, CA. Human Factors Society.

[Hoenig and Heisey, 2001] Hoenig, J. M. and Heisey, D. M. (2001). The abuse of power: The pervasive fallacy of power calculations for data analysis. *The American Journal of Psychology*, 55:19–24.

[Hollan et al., 2000] Hollan, J., Hutchins, E., and Kirsh, D. (2000). Distributed cognition: Toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 7(2):174–196.

[Hutchins, 1995] Hutchins, E. (1995). *Cognition in the Wild*. MIT Press.

[Huynh et al., 2002] Huynh, D., Karger, D., and Quan, D. (2002). Haystack: A platform for creating, organizing and visualizing information using RDF. In *Semantic Web Workshop*, Hawaii, USA.

[International Standards Organization, 2008] International Standards Organization (2008). ISO 9241: Ergonomics of human-system interaction.

[Iqbal et al., 2005] Iqbal, S. T., Adamczyk, P. D., Zheng, X. S., and Bailey, B. P. (2005). Towards an index of opportunity: understanding changes in mental workload during task execution. In *CHI '05: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 311–320, New York, NY, USA. ACM Press.

[Iqbal and Bailey, 2005] Iqbal, S. T. and Bailey, B. P. (2005). Investigating the effectiveness of mental workload as a predictor of opportune moments for interruption. In *CHI '05: Extended Abstracts on Human Factors in Computing Systems*, pages 1489–1492, New York, NY, USA. ACM.

[Johanson et al., 2001] Johanson, B., Ponnekanti, S., Sengupta, C., and Fox, A. (2001). Multibrowsing: Moving web content across multiple displays. In *UbiComp '01: Proceedings of the 3rd International Conference on Ubiquitous Computing*, pages 346–353, London, UK. Springer-Verlag.

[Jones, 2008] Jones, W. (2008). How is information personal? In *3rd Invitational Workshop on Personal Information Managementat CHI 2008: The Disappearing Desktop (PIM 2008)*.

[Jones et al., 2001] Jones, W., Bruce, H., and Dumais, S. (2001). Keeping found things found on the web. In *CIKM '01: Proceedings of the Tenth International Conference on Information and Knowledge Management*, pages 119–126, New York, NY, USA. ACM Press.

[Jones et al., 2002] Jones, W., Dumais, S., and Bruce, H. (2002). Once Found, What Then? A Study of 'Keeping' Behaviors in the Personal Use of Web Information. In *Proceedings of the American Society for Information Science and Technology*, volume 39, pages 391–402. American Society for Information Science and Technology.

[Jones et al., 2008] Jones, W., Klasnja, P., Civan, A., and Adcock, M. (2008). The personal project planner: Planning to organize personal information. In *CHI 2008 Proceedings of the ACM Conference on Human Factors in Computing Systems*, page 681.

[Jones and Teevan, 2007] Jones, W. and Teevan, J. (2007). *Personal Information Management*. University of Washington Press, Seattle, Washington.

[Jordan, 2000] Jordan, P. W. (2000). *Designing Pleasurable Products: An Introduction to the New Human Factors*. CRC.

[Kaasten et al., 2002] Kaasten, S., Greenberg, S., and Edwards, C. (2002). How people recognize previously seen www pages from titles, urls and thumbnails. In Faulkner, X., Finlay, J., and Detienne, F., editors, *People and Computers*, volume XVI of *Proceedings of Human Computer Interaction, BCS Conference Series*, pages 247–265. Springer Verlag.

[Kahneman, 1973] Kahneman, D. (1973). *Attention and Effort*. Prentice-Hall, Englewood Cliffs, NJ.

[Karger and Jones, 2006] Karger, D. and Jones, W. (2006). Data unification in personal information management. *Communications of the Association for Computing Machinery (CACM)*, 49(1):77–82.

[Karger and Quan, 2004] Karger, D. and Quan, D. (2004). Haystack: a user interface for creating, browsing, and organizing arbitrary semistructured information. In *CHI '04: Extended Abstracts on Human Factors in Computing Systems*, pages 777–778, New York, NY, USA. ACM Press.

[Kelley and Chapanis, 1982] Kelley, J. and Chapanis, A. (1982). How professional persons keep their calendars: Implications for computerization. *The British Psychological Society*.

[Kelly, 2006] Kelly, D. (2006). Evaluating personal information management behaviors and tools. *Communications of the Association for Computing Machinery (CACM)*, 49(1):84–86.

[Kelly and Teevan, 2003] Kelly, D. and Teevan, J. (2003). Implicit feedback for inferring user preference: a bibliography. *SIGIR Forum*, 37(2):18–28.

[Kelly and Teevan, 2007] Kelly, D. and Teevan, J. (2007). *Understanding What Works: Evaluating PIM Tools*, chapter 12, page 17. University of Washington Press, Seattle, Washington.

[Kincaid et al., 1985] Kincaid, C. M., Dupont, P. B., and Kaye, A. R. (1985). Electronic calendars in the office: an assessment of user needs and current technology. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 3(1):89–102.

[Kirsh, 2006] Kirsh, D. (2006). Personal information objects and burden of multiple personal spaces. In *Proceedings of the 2nd Invitational Workshop on Personal Information Management at SIGIR 2006*.

[Klingner et al., 2008] Klingner, J., Kumar, R., and Hanrahan, P. (2008). Measuring the task-evoked pupillary response with a remote eye tracker. In *Eye Tracking Research and Applications Symposium*, Savannah, Georgia.

[Kok, 1997] Kok, A. (1997). Event-related-potential (ERP) reflections of mental resources: a review and synthesis. *Biological Psychology*, 45(1):19–56.

[Komninos et al., 2008] Komninos, A., Barrie, P., and Baillie, L. (2008). Holistic PIM: Managing personal information for a nomadic generation. In *3rd Invitational Workshop on Personal Information Managementat CHI 2008: The Disappearing Desktop (PIM 2008)*.

[Kramer, 1991] Kramer, A. F. (1991). *Physiological metrics of mental workload: A review of recent progress*, pages 279–328. CRC Press.

[Krippendorff, 2004] Krippendorff, K. (2004). *Content analysis: An introduction to its methodology*. Sage Publications, Inc.

[Kwasnik, 1989] Kwasnik, B. (1989). How a personal document's intended use or purpose affects its classification in an office. *SIGIR Forum*, 23(SI):207–210.

[Lansdale, 1988] Lansdale, M. W. (1988). The psychology of personal information management. *Applied Ergonomics*, 19:55–66.

[Layard, 2006] Layard, R. (2006). *Happiness*. Penguin.

[Lenth, 2001] Lenth, R. V. (2001). Some practical guidelines for effective sample size determination. *The American Statistician*, 55(3):187–193.

[Lerman, 1996] Lerman, J. (1996). Study design in clinical research: sample size estimation and power analysis. *Canadian Journal of Anaesthesia*, 43(2):184–191.

[Levy, 2005] Levy, D. M. (2005). To grow in wisdom: Vannevar Bush, Information Overload, and the Life of Leisure. In *JCDL '05: Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL)*, pages 281–286, New York, NY, USA. ACM Press.

[Mackay, 1988] Mackay, W. E. (1988). Diversity in the use of electronic mail: a preliminary inquiry. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 6(4):380–397.

[Malone, 1983] Malone, T. W. (1983). How do people organize their desks?: Implications for the design of office information systems. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 1(1):99–112.

[Marshall, 2002] Marshall, S. (2002). The index of cognitive activity: measuring cognitive workload. *Proceedings of the 2002 IEEE 7th Conference on Human Factors and Power Plants, 2002.*, pages 75–79.

[Maslow, 1943] Maslow, A. H. (1943). A theory of human motivation. *Psychological Review*, 50:370–396.

[Microsoft, 2006] Microsoft (2006). Windows desktop search. http://www.microsoft.com/windows/products/winfamily/desktopsearch/.

[Mills and Scholtz, 2001] Mills, K. and Scholtz, J. (2001). *Situated Computing: The Next Frontier for HCI Research*. Addison-Wesley Professional.

[Miyata and Norman, 1986] Miyata, Y. and Norman, D. A. (1986). *The Control of Multiple Activities*. Lawrence Erlbaum Associates, Hillsdale, NJ.

[Mori et al., 2003] Mori, G., Paternò, F., and Santoro, C. (2003). Tool support for designing nomadic applications. In *IUI '03: Proceedings of the 8th International Conference on Intelligent User Interfaces*, pages 141–148, New York, NY, USA. ACM Press.

[Muckler and Seven, 1992] Muckler, F. A. and Seven, S. A. (1992). Selecting performance measures: "objective" versus "subjective" measurement. *Human Factors*, 34(4):441–455.

[Nardi and O'Day, 2000] Nardi, B. A. and O'Day, V. (2000). *Information Ecologies: Using Technology with Heart*. MIT Press.

[Nardi et al., 2000] Nardi, B. A., Whittaker, S., and Bradner, E. (2000). Interaction and outeraction: Instant messaging in action. In *CSCW '00: Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work*, pages 79–88, New York, NY, USA. ACM.

[Nelson, 1994] Nelson, M. R. (1994). We have the information you want, but getting it will cost you!: Held hostage by information overload. *Crossroads*, 1(1):11–15.

[Nielsen, 2003] Nielsen, J. (2003). IM, not IP (Information Pollution). *ACM Queue*, 1(8):76–75.

[Nisbett and Wilson, 1977] Nisbett, R. E. and Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84(3):231–259.

[Norman, 1988] Norman, D. A. (1988). *The Psychology Of Everyday Things*. Basic Books.

[Norman, 1999] Norman, D. A. (1999). *The Invisible Computer: Why Good Products Can Fail, the Personal Computer Is So Complex, and Information Appliances Are the Solution*. The MIT Press.

[Norman, 2003] Norman, D. A. (2003). *Emotional Design: Why We Love (or Hate) Everyday Things*. Basic Books.

[O'Donnell and Eggemeier, 1986] O'Donnell, R. D. and Eggemeier, F. T. (1986). *Workload assessment methodology*, chapter 2, pages 42/1–42/49. Handbook of perception and human performance: Vol. 2. Cognitive processes and performance. Wiley, New York.

[Oquist et al., 2004] Oquist, G., Goldstein, M., and Chincholle, D. (2004). Assessing usability across multiple user interfaces. In Seffah, A. and Javahery, H., editors, *Multiple User Interfaces: Cross-Platform Applications and Context-Aware Interfaces*, pages 227–349. John Wiley & Sons.

[Oulasvirta et al., 2005] Oulasvirta, A., Tamminen, S., Roto, V., and Kuorelahti, J. (2005). Interaction in 4-second bursts: The fragmented nature of attentional resources in Mobile HCI. In *CHI '05: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 919–928, New York, NY, USA. ACM Press.

[Park et al., 2006] Park, S., Harada, A., and Igarashi, H. (2006). Influences of personal preference on product usability. In *CHI '06: Extended Abstracts on Human Factors in Computing Systems*, pages 87–92, New York, NY, USA. ACM.

[Payne, 1993] Payne, S. J. (1993). Understanding calendar use. *Human-Computer Interaction*, 8(2):83–100.

[Pérez-Quiñones et al., 2008] Pérez-Quiñones, M., Tungare, M., Pyla, P., and Harrison, S. (2008). Personal information ecosystems: Design concerns for net-enabled devices. In *Proceedings of the VI Latin American Web Congress - LA-Web 2008*.

[Perry et al., 2001] Perry, M., O'Hara, K., Sellen, A., Brown, B., and Harper, R. (2001). Dealing with mobility: Understanding access anytime, anywhere. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 8(4):323–347.

[Pyla et al., 2009] Pyla, P., Tungare, M., Holman, J., and Pérez-Quiñones, M. (2009). Continuous user interfaces for seamless task migration. In *Proceedings of the 13th International Conference on Human-Computer Interaction, HCII 2009*.

[Pyla et al., 2006] Pyla, P., Tungare, M., and Pérez-Quiñones, M. (2006). Multiple user interfaces: Why consistency is not everything, and seamless task migration is key. In *Proceedings of the CHI 2006 Workshop on The Many Faces of Consistency in Cross-Platform Design*.

[R Development Core Team, 2008] R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

[Rath et al., 2008] Rath, A. S., Weber, N., Kröll, M., Granitzer, M., Dietzel, O., and Lindstaedt, S. N. (2008). Context-aware knowledge services. In *3rd Invitational Workshop on Personal Information Managementat CHI 2008: The Disappearing Desktop (PIM 2008)*.

[Reid et al., 1982] Reid, G. B., Eggemeier, F. T., and Shingledecker, C. A. (1982). Subjective Workload Assessment Technique. Technical report, Air Force Flight Test Center, Edwards, CA.

[Reid and Nygren, 1988] Reid, G. B. and Nygren, T. E. (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload. *Human mental workload*, 185:218.

[Rekimoto, 1997] Rekimoto, J. (1997). Pick-and-drop: A direct manipulation technique for multiple computer environments. In *UIST '97: Proceedings of the 10th Annual ACM Symposium on User Interface Software and Technology*, pages 31–39, New York, NY, USA. ACM.

[Richter, 2005] Richter, K. (2005). A Transformation Strategy for Multi-device Menus and Toolbars. In *CHI '05: Extended Abstracts on Human Factors in Computing Systems*, pages 1741–1744, New York, NY, USA. ACM Press.

[Robbins, 2008] Robbins, D. C. (2008). TapGlance: Designing a unified smartphone interface for Personal Information Management. In *3rd Invitational Workshop on Personal Information Managementat CHI 2008: The Disappearing Desktop (PIM 2008)*.

[Roscoe, 1984] Roscoe, A. H. (1984). Assessing pilot workload in flight. In *AGARD Flight Test Techniques*, number N 84-34396 24-01, pages 1–32.

[Rouse et al., 1993] Rouse, W., Edwards, S., and Hammer, J. (1993). Modeling the dynamics of mental workload and human performance in complex systems. *Systems, Man and Cybernetics, IEEE Transactions on*, 23(6):1662–1671.

[Rubio et al., 2004] Rubio, S., Díaz, E., Martín, J., and Puente, J. M. (2004). Evaluation of subjective mental workload: A comparison of SWAT, NASA TLX, and Workload Profile methods. *Applied Psychology: An International Review*, 53(1):61–86.

[Savitzky and Golay, 1964] Savitzky, A. and Golay, M. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36(8):1627–1639.

[Schacter, 1977] Schacter, D. L. (1977). EEG theta waves and psychological phenomena: A review and analysis. *Biological Psychology*, 5(1):47–82.

[Schick et al., 1990] Schick, A. G., Gordon, L. A., and Haka, S. (1990). Information overload: A temporal approach. *Accounting, Organizations and Society*, 15(3):199–220.

[Schryver, 1994] Schryver, J. C. (1994). Experimental validation of navigation workload metrics. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, 38:340–344(5).

[Schultheis and Jameson, 2004] Schultheis, H. and Jameson, A. (2004). Assessing cognitive load in adaptive hypermedia systems: Physiological and behavioral methods. *Adaptive Hypermedia and Adaptive Web-Based Systems*, pages 225–234.

[Shenk, 1998] Shenk, D. (1998). *Data Smog: Surviving the Information Glut Revised and Updated Edition*. HarperOne.

[Singh, 2006] Singh, G. (2006). PIM for Mobility. In *Proceedings of the 2nd Invitational Workshop on Personal Information Management at SIGIR 2006.*

[Speyer et al., 1988] Speyer, J. J., Fort, A., Fouillot, J., and Blomberg, R. D. (1988). *Dynamic methods for assessing workload for minimum crew certification*, pages 316–88. Number IB 316-88-06.

[Spinuzzi, 2001] Spinuzzi, C. (2001). Grappling with distributed usability: A cultural-historical examination of documentation genres over four decades. *Journal of Technical Writing and Communication*, 31(1):41–59.

[Tauscher and Greenberg, 1997] Tauscher, L. and Greenberg, S. (1997). How people revisit web pages: empirical findings and implications for the design of history systems. *International Journal of Human-Computer Interaction*, 47(1):97–137.

[Teevan et al., 2004] Teevan, J., Alvarado, C., Ackerman, M. S., and Karger, D. (2004). The perfect search engine is not enough: A study of orienteering behavior in directed search. In *CHI '04: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 415–422, New York, NY, USA. ACM Press.

[Teevan et al., 2007] Teevan, J., Capra, R. G., and Pérez-Quiñones, M. (2007). *How people find information*, chapter 3, page 17. University of Washington Press, Seattle, Washington.

[Teevan and Jones, 2008] Teevan, J. and Jones, W. (2008). PIM 2008: Personal Information Management: The Disappearing Desktop, a CHI 2008 Workshop. Personal discussions with workshop participants.

[Thevenin and Coutaz, 1999] Thevenin, D. and Coutaz, J. (1999). Plasticity of user interfaces: Framework and research agenda. In *Interact*, pages 110–117, Edinburgh. IFIP.

[Tsang and Velazquez, 1996] Tsang, P. and Velazquez, V. (1996). Diagnosticity and multidimensional subjective workload ratings. *Ergonomics*, 39(3):358–381.

[Tungare and Pérez-Quiñones, 2008a] Tungare, M. and Pérez-Quiñones, M. (2008a). An exploratory study of personal calendar use. Technical report, Computing Research Repository (CoRR).

[Tungare and Pérez-Quiñones, 2008b] Tungare, M. and Pérez-Quiñones, M. (2008b). It's not what you have, but how you use it: Compromises in mobile device use. Technical report, Computing Research Repository (CoRR).

[Tungare and Pérez-Quiñones, 2008c] Tungare, M. and Pérez-Quiñones, M. (2008c). Thinking outside the (beige) box: Personal information management beyond the desktop. In *Proceedings of the 3rd Invitational Workshop on Personal Information Management, PIM 2008, a CHI 2008 workshop*.

[Tungare et al., 2006] Tungare, M., Pyla, P., Sampat, M., and Pérez-Quiñones, M. (2006). Defragmenting information using the Syncables framework. In *Proceedings of the 2nd Invitational Workshop on Personal Information Management at SIGIR 2006*.

[Tungare et al., 2007] Tungare, M., Pyla, P., Sampat, M., and Pérez-Quiñones, M. (2007). Syncables: A framework to support seamless data migration across multiple platforms. In *IEEE International Conference on Portable Information Devices (IEEE Portable)*.

[Weiser, 1991] Weiser, M. (1991). The computer for the 21st century. *Scientific American*, 265(3):66–75.

[Weiser, 1994] Weiser, M. (1994). The world is not a desktop. *Interactions*, 1(1):7–8.

[Whittaker and Hirschberg, 2001] Whittaker, S. and Hirschberg, J. (2001). The character, value, and management of personal paper archives. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 8(2):150–170.

[Whittaker et al., 2002a] Whittaker, S., Jones, Q., and Terveen, L. (2002a). Contact management: identifying contacts to support long-term communication. In *CSCW '02: Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work*, pages 216–225, New York, NY, USA. ACM.

[Whittaker et al., 2002b] Whittaker, S., Jones, Q., and Terveen, L. (2002b). Managing long term communications: conversation and contact management. *Proceedings of the 35th Annual Hawaii International Conference on System Sciences, 2002.*, pages 1070–1079.

[Whittaker and Sidner, 1996] Whittaker, S. and Sidner, C. (1996). Email overload: exploring personal information management of email. In *CHI '96: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 276–283, New York, NY, USA. ACM Press.

[Whittaker et al., 2000] Whittaker, S., Terveen, L., and Nardi, B. A. (2000). Let's stop pushing the envelope and start addressing it: A reference task agenda for HCI. *Human-Computer Interaction*, 15(2):75–106.

[Wickens, 1992] Wickens, C. (1992). *Engineering Psychology and Human Performance*. HarperCollins Publishers, New York.

[Wierwille and Casali, 1983] Wierwille, W. W. and Casali, J. G. (1983). A validated rating scale for global mental workload measurement application. In *Proceedings of the Human Factors Society 27th Annual Meeting*, pages 129–133, Santa Monica, CA. Human Factors Society.

[Wierwille et al., 1985] Wierwille, W. W., Rahimi, M., and Casali, J. G. (October 1985). Evaluation of 16 measures of mental workload using a simulated flight task emphasizing mediational activity. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 27:489–502.

[Wilson and Eggemeier, 1991] Wilson, G. F. and Eggemeier, F. T. (1991). *Psychophysiological assessment of workload in multi-task environments*, chapter 12. Taylor and Francis.

[Wilson and Eggemeier, 2006] Wilson, G. F. and Eggemeier, F. T. (2006). *Mental Workload Measurement*, pages 814–817. International Encyclopedia of Ergonomics and Human Factors. CRC Press.

[Woerndl and Woehrl, 2008] Woerndl, W. and Woehrl, M. (2008). SeMoDesk: Towards a mobile semantic desktop. In *3rd Invitational Workshop on Personal Information Managementat CHI 2008: The Disappearing Desktop (PIM 2008)*.

Author Index