Using Visual Histories to Reconstruct the Mental Context of Suspended Activities

Adam Rule, Aurélien Tabard, Jim Hollan

HAL Id: hal-01493400
https://hal.science/hal-01493400v1

Using Visual Histories to Reconstruct the Mental Context of Suspended Activities

Adam Rule[1], Aurélien Tabard[2], Jim Hollan[1]

[1]University of California, San Diego,

9500 Gillman Drive,

La Jolla, CA 92093

[2] Université de Lyon, CNRS,

Université Lyon 1, LIRIS UMR5205,

F-69622, France

Authors' Mini-bios:

Adam Rule (acrule@ucsd.edu, adamrule.com) is a cognitive scientist with an interest in interruption and multitasking; he is a PhD Candidate in the Department of Cognitive Science of UC San Diego.

Aurélien Tabard (aurelien.tabard@univ-lyon1.fr, www.tabard.fr/) is a Computer Scientist with an interest in personal informatics, activity based computing and information visualization; he is an Assistant Professor in the Computer Science department at Université Lyon 1, LIRIS.

Jim Hollan (hollan@ucsd.edu, hci.ucsd.edu/hollan/) is a cognitive scientist with interests in human-computer interaction and information visualization; he is Professor of Cognitive Science and Computer Science & Engineering at UC San Diego and founding co-director of the Design Lab.

Using Visual Histories to Reconstruct the Mental Context of Suspended Activities

Running Head: Using Visual Histories to Reconstruct Mental Context

Abstract

Resuming suspended activities is fundamental to knowledge work. As activities grow in complexity or fade from memory, they become increasingly difficult to cognitively resume. Motivated by the efficiency of images for cueing autobiographical memory, we conducted two studies of how visual histories of computer-mediated activity might aid users in restoring the mental context of suspended activities. In the first exploratory study we demonstrate that users provided a visual history consisting of small thumbnail images of their desktop can reconstruct detailed mental contexts of earlier computer work. In the second observational study we examine methods and cues that web developers, scientific writers, and graphic designers currently use to restore mental context. We find that participants from all three professions engage in *situated sensemaking*, reconstructing context by generating stories to explain cues in their documents with implicit meaning, such as sudden changes in writing quality or the shape of blocks of code. We identify characteristics of each activity that influence the information, artifacts, and strategies participants used to reconstruct context. We conclude that visual histories can help users reconstruct mental context, but that challenges remain in designing lightweight image-based cues that help users reconstruct context for a range of activities.

contents

1. INTRODUCTION

Significant projects are rarely finished in one sitting but are often suspended and resumed multiple times before they are complete. This can be a consequence of working on multiple projects at the same time, or having projects span days, weeks, or months. The resulting fragmentation of work may have little impact on simple procedural activities; resuming a form-filling activity after a break may be as simple as finding the next empty field. But knowledge work often involves complex activities in which it is harder to tell where one left off and what one intended to do next. These complex activities involve constructing and manipulating fragile mental states. Such states motivate and guide the activity, but are ephemeral and fade from memory over time, and it can be difficult to resume the activity without first reconstructing this mental context. Take, for example, the programmer in Figure 1 who, after being interrupted by a colleague, needs to reconstruct the complex program flow before resuming their development work. A key challenge for human-computer interaction is helping people restore mental contexts associated with ongoing activities after short interruptions, longer breaks, and other activities push them from memory.

(Figure 1 about here)

Resuming suspended activities is a well-known problem. There is a long tradition of writers and academics creating workshops and retreats where they can avoid interruptions and focus for extended periods of time (Newport, 2015; Pollan, 2008). This seclusion and undivided attention is not possible for many of today's knowledge workers who struggle to focus for more than a dozen or so minutes at a time (Gonzalez & Mark, 2004) as they manage multiple projects while responding to a steady stream of emails, messages, and alerts. Because they regularly switch between complex intellectual activities, knowledge workers often need to restore mental context. Unfortunately, current software systems provide little support for doing so, even as they are increasingly a source of disruption.

Designing effective resumption aids has been a key challenge for human-computer interaction since the start of personal computing. Writing about computer support for multitasking in 1986, Mitaya and Norman claimed that:

> "…sufficient information should be saved with the suspended task so that when the activity is resumed, it can be continued where it left off (recovering the active thoughts is the hard part)." (p. 276)

Research and development in the intervening decades has provided little guidance for "the hard part" of helping users recover "active thoughts". Instead, computer systems have slowly improved their ability to help users restore documents and workspaces. Moreover, memory research has acknowledged the richness of mental context for ongoing activities but primarily measured *how many* events people remember when using various aids, not *how detailed* their memories for these events are.

We believe one promising way to help people restore rich mental contexts lies in cuing their memory with visual histories of past work. Images from one's past are

excellent cues of autobiographical memory because they resurface features present at the time of memory formation. And while computers cannot track what users are thinking, they can track what they may be seeing. We believe visual histories provided by images of past work may help users not only remember suspended activities, but also cue reconstruction of the complex network of thoughts they had while performing them, easing resumption of those activities.

Our aim in this research is to help people restore mental context by graphically showing computer activity spanning multiple applications. This vision differs from most prior activity history research in goal, approach, and scope. Visualizations of computer activity have been used for a variety of purposes including helping people re-find documents, generate tutorials, or retrace a set of operations (Grabler, Agrawala, Li, Dontcheva, & Igarashi, 2009; Hailpern et al., 20011; Heer, Mackinlay, Stolte, & Agrawala 2008). However, few visualizations have been explicitly designed to help people remember what they were *thinking* in the past. Many visualizations of computer activity are abstract and list operations performed or documents accessed. In contrast, we propose a highly concrete and graphical approach, presenting images of computer activity. Lastly, most computer histories are limited to a single document or application. But everyday computer activity routinely spans multiple applications and we aim to represent that cross-application activity.

Capturing and representing a visual history of past computer activity confronts multiple challenges. There are privacy concerns with recording and storing images of potentially sensitive work. Recorded images may not be distinct enough to prompt memories about specific moments in time, but only general ways of working. Different visual cues may be needed to resume different types of work. Large collections of images or videos documenting ones' past work may be overwhelming or tedious to review. Fully addressing these challenges is beyond the scope of this paper. Instead, the studies presented here address more fundamental questions about the nature of mental context and the feasibility of reconstructing it with visual histories.

In the first study, we explore whether viewing small thumbnail images of prior desktop arrangements can help knowledge workers reconstruct mental context. Our results show that people can reconstruct what they were doing *and thinking* in the past while viewing thumbnail images of prior computer activity. In some cases, they are able to reconstruct surprisingly detailed context from relatively sparse cues (Figure 5). We also find that, on average, images of the participant's entire computer screen cued reconstruction of more detailed and accurate context than animated thumbnails, or thumbnails showing only part of the participant's screen.

While our thumbnail study provides evidence for the feasibility of using visual histories to cue reconstruction of mental context, it raises additional questions about the nature of mental context and how to design more effective cues. For example: What types of mental context do knowledge workers typically reconstruct when resuming suspended activities? What visualizations of past activity might best represent an activity and cue the most detailed and accurate context? How might these visualizations be tailored to the particular activity represented? How might visual histories augment knowledge workers'

current methods of activity resumption?

These questions led us to conduct a second study of the practices and cues knowledge workers currently use to restore mental context for a range of activities. We recorded and interviewed ten web developers, scientific writers, and graphic designers as they resumed long-term activities during their everyday work. Our results show that all three types of knowledge worker engage in *situated sensemaking*, reconstructing mental context by generating stories to explain subtle, implicit cues in their documents such as changes in writing quality or the shape of blocks of code. We find that they use these cues not only to remember where they left off, but also their motivations, intentions, and plans. From our observations we highlight the cognitive and collaborative characteristics of each activity that seemed to systematically influence the cues, artifacts, and strategies participants used to reconstruct context.

We conclude that visual histories can help users reconstruct the mental context of a wide range of computer-mediated activities, and outline challenges in designing consistently lightweight and effective image-based cues for a range of activities. Before describing the studies and their results, we review related literature on fragmented work and computer support for resuming suspended activities.

2. RELATED WORK

2.1. Fragmented Work

Knowledge workers routinely manage multiple projects while keeping tabs on multiple steams of communication. Doing so requires interleaving and overlapping activities. In their seminal workplace study, Gonzalez and Mark found that information workers handle 10 distinct projects in a typical day, spending an average of just 12 minutes on each one before being interrupted, or interrupting themselves, to work on another (Gonzalez & Mark, 2004). That same year, a diary study found that information workers report spending, on average, less than an hour on each activity before switching (Czerwinski, Horvitz, & Wilhite, 2004). Suspended activities may not be resumed for some time. One study found that, when interrupted, information workers spend an average of 25 minutes on others activities before returning to the suspended one, and that more than a quarter of suspended activities are not resumed until the next day, or later (Mark, González & Harris, 2005).

These measures were taken more than a decade ago and reflect the state of knowledge work before the widespread adoption of smartphones and social media. Fragmentation is likely even more pronounced today, as demonstrated by one study of college students which found they switched between computer application windows on average once every 48 seconds (Mark, Wang, & Niiya, 2014). While some of this fragmentation is caused by self-interruptions (Dabbish, Mark, & González, 2011) and scheduled events like meetings, much can be attributed to using information technology. When 13 knowledge workers gave up email for a week, they focused for twice as long on each application window before switching, though there was significant variation between participant and type of activity (Mark, Voida, & Cardello, 2012).

2.2. Consequences of Fragmentation

Interruptions are not necessarily harmful and are often required for collaboration. By one estimate, information workers benefit from nearly two-thirds (64%) of interruptions, for example, by gaining new information (O'Conaill & Frohlich, 1995). However, there is growing evidence that working in jumps and starts can take more time, produce lower quality work, and increase stress when compared to working without interruption.

Being interrupted while programming increases the overall time needed to finish the programming activity. From their observations, van Solingen, Berghout, and van Latum (1998) estimated that programmers take an average of 15 minutes to recover from each interruption and spend over an hour each day managing them. Parnin and Deline (2010) similarly concluded "when resuming interrupted work, developers experience increased time to perform the task" as well as "increased errors, increased loss of knowledge, and increased forgetting to perform critical tasks".

Writers have been shown to produce poorer quality work when interrupted. A study of student writers found they produced poorer quality essays when they were interrupted, regardless of whether the interruptions occurred while they were outlining their essay or writing the main text (Foroughi, Werner, Nelson, & Boehm-Davis, 2014). However, this effect is complex. Lottridge et al. (2015) found that heavy multi-taskers wrote substantially worse essays when interrupted with irrelevant messages, but *substantially better* essays when interrupted with relevant information. These were even better than essays written by light multi-taskers who received the same helpful information. The impact of interruptions then depends not only on the interruption's relevance, but also on the person being interrupted.

Working in short segments can also contribute to stress. In the diary study of interruption mentioned above, participants found it harder to restart suspended tasks than to start new ones (Czerwinski, Horvitz, & Wilhite, 2004). Recent work by Mark, Iqbal, Czerwinski and Johns (2015) has correlated more time in email, face-to-face conversations, and more task-switches with feeling less productive.

It is difficult to generalize about phenomena that are likely sensitive to context, the nature of specific tasks, and individual differences. However, the evidence to date can be summarized as follows: fragmentation pervades knowledge work, information technology seems to contribute to this fragmentation, frequent interruption can result in lower quality, longer duration, and more error prone work for activities such as programming and writing, and pervasive task switching can contribute to feelings of stress and lower productivity.

2.3. Cognitive Challenges

Part of the challenge with frequent interruption is remembering to resume the suspended activity. It is not uncommon for knowledge workers to forget an interrupted activity altogether (O'Conaill & Frohlich, 1995). This is known as a failure of prospective memory (Ellis & Kvavilashvili, 2000).

Another challenge is recovering the physical or digital resources needed to continue the activity. For computer-based activities this includes relevant documents and files, which can be difficult to find because they are obscurely named, filed deep in a folder structure, or simply hidden behind the windows of intervening tasks. Much work has gone into making documents and files easier to find, but, even if they are open, simply having part or all of a window obscured can make the associated activity harder to remember and resume (Iqbal & Horvitz, 2007; Jo, Kim, & Seo, 2015; Trafton, Altmann, & Brock 2005). As a result, it is common for knowledge workers to search for suspended activities by cycling through their open windows (Parnin & Rugaber, 2011; Iqbal & Horvitz, 2007). However, as Iqbal and Horvitz (2007) noted, "even when the user finally returned to the suspended application, substantial time appeared to be devoted to resuming the task state and, presumably, the mental state that they had been in prior to suspension". The challenge with fragmented work is not just remembering the suspended task and finding the right documents. *People need to restore mental context as well*.

One widely held psychological model of interruption proposes that when an interruption occurs, the goal of the interrupted activity slowly fades from memory as the goal for the interrupting activity become more active (Altmann & Trafton, 2002). According to this Memory for Goals model, the challenge with resuming the interrupted activity is activating the suspended goal by searching through memory or external cues. Beyond the availability of a reorienting cue, such as a cursor placed over the next field to edit in a form, controlled studies have found that the time it takes to remember and resume a suspended activity is a function of the duration of the interruption (Altmann & Trafton, 2002; Monk, Trafton, & Boehm-Davis, 2008), its similarity to the interrupted task (Gillie & Broadbent, 1989), the difficulty of the interrupting and interrupted tasks (Monk, Trafton, & Boehm-Davis, 2008), and an individual's experience with the interrupted task (Hess & Detweiler, 1994).

The Memory for Goals model provides a useful description of how memories for interrupted activities fade over time and are obscured by memories for intervening activities. However, as we discuss next, assembling the mental context needed to resume an activity is more complex than simply remembering a suspended goal. Activities such as writing and programming involve rich mental context about the state of the activity and one's intentions, which also need to be restored. Psychologists have begun to recognize and account for this complexity in computational models of activity resumption (Salvucci, 2010).

2.4. Restoring Mental Context

Our current understanding of what mental context is and how people reconstruct it comes mainly from observing programmers. Programming activities are notoriously difficult to resume as they involve creating and manipulating complex mental states that fade quickly during an interruption. Recreating these states can be difficult. Many studies have looked at how people resume programming activities (Safer & Murphy, 2007; van Solingen, Berghout, & van Latum, 1998), but Chris Parnin's surveys and analyses of in-the-wild programming activity stand out for their detail and cognitive focus (Parnin & Deline, 2010; Parnin & Rugaber, 2011, Parnin & Rugaber, 2012). Parnin and his

collaborators found that programmers routinely use both suspension strategies (e.g., note-taking and cue-priming) and resumption strategies (e.g., cue-seeking and systematic review) that involve a range of information sources (e.g., program files, notes, run-time information, compile errors, and task lists) to preserve mental context across interruptions. Programmers proactively leave in-line comments, intentionally write code that will throw compiler errors so they will be returned to a specific section of code when resuming, and retrospectively inspect their code and console logs for reminders. By documenting the specific challenges interruptions pose for programmers and the strategies they use to overcome them, Parnin's work provides concrete guidance for the development of programming aids such as integrated development environments (IDEs). Similar observations motivated the development of the Mylyn extension for the Eclipse IDE, which filters the items shown in Eclipse's file tree based on the programmer's declaration of their current task (Kersten, Elves, & Murphy, 2006). Such advancements are especially valuable at a time when an increasing number of knowledge workers program as part of their everyday work.

However, the highly structured nature of programming activities has encouraged researchers to characterize restoring mental context as a well-defined information processing problem. Parnin's work focuses on how programmers restore suspended goals, plans, and actions and describes "context restoration" as a step after goal restoration wherein the programmer recalls which lines of code they were editing to achieve that particular goal (Parnin & Rugaber, 2011). But as humorously indicated in Figure 1, even for programming activities, mental context is richer than a list of goals, plans, actions, and files. It includes a host of ephemeral thoughts about the activity, some of which are not goal-directed. Moreover, mental context that supports non-programming activities may have drastically different content and structure than researchers have observed to date.

A final limitation of prior work is a focus on explicit cues, such as handwritten notes, but not implicit cues, such as the shape of a block code. These implicit cues are often a natural byproduct of performing the activity but may carry meaning for those who originally performed the activity. For example, simply seeing a block of code that is shorter or less ragged than expected may prompt the viewer to recall that they had tried implementing a complex algorithm, but later abandoned it for a simpler one. Our second study identifies some of these implicit cues by observing the information, artifacts, and strategies three types of knowledge worker use to restore mental context for various activities.

2.5. Aiding Resumption

While there are still significant gaps in our knowledge of mental context and how people restore it, many studies have nonetheless demonstrated how computation can help people manage fragmented work and resume suspended activities. Building on psychological models of interruption (Trafton, Altmann, Brock, & Mintz, 2003), we classify these interventions according to three moments in the interruption cycle where they intervene: before suspension, upon suspension, and upon resumption of the primary activity.

Before suspension, software can intercept incoming interruptions and manage them in a variety of ways (McFarlane & Latorella, 2002). Extensive work has gone into designing interruption management systems (Hincapié-Ramos, Voida, & Mark, 2011) with several studies quantifying worker's interruptibility by measuring their general computer activity (Horvitz, Koch, & Apacible, 2004), physical office activity (Fogarty, Hudson & Lai, 2004), or the characteristics of their current and interrupting activities (Iqbal and Bailey, 2006). While these studies have modeled bad times to interrupt or suspend an activity, other research has revealed that the best times to interrupt seem to be points of low cognitive demand, which often occur at breakpoints between components of an activity (Iqbal & Bailey, 2007; Monk, Boehm-Davis, & Trafton, 2002). Messages delivered at breakpoints are better received and lead to faster resumption of the interrupted task than messages received in the middle of a complex sequence of actions (Adamczyk & Bailey, 2004; Ho & Intille, 2005). Beyond recognizing breakpoints as good times to interrupt, researchers have developed breakpoint detectors that automatically defer interruptive messages and alerts to these times (Iqbal & Bailey, 2007). Even without assistance computer users routinely manage interruptions such as email alerts by ignoring them until the next breakpoint (Salvucci, 2010). Rescheduling unnecessary and costly interruptions is a worthwhile approach to managing fragmented work. However, many interruptions are urgent, occur face-to-face, or are self-initiated. In these cases computers probably should not, or simply cannot, intercept them. Instead, they might help people prepare for the impending suspension and later resumption.

Upon suspension, and more specifically during the lag between deciding to suspend the current activity and attending to the interrupting one, systems can encourage people to get to a good stopping point before switching activities or to leave cues that may help them reorient to the interrupted activity later. Trafton, Altmann, Brock & Mintz (2003) noted in one laboratory study that participants resumed an interrupted gaming task faster if they were given a longer lag between stopping one task and starting the next one in which they could mentally rehearse what they needed to do when they returned. Similarly, computer users often rearrange their windows or jot quick notes to remind themselves what they need to do when they return to the current activity (Parnin & Deline, 2010). To our knowledge, no human-computer interaction research has explicitly targeted the moment of suspension by encouraging computer users to pause and think about what they need to do when they return, or to proactively leave hints and notes for their future selves. This may be a productive line of research, but in many cases interruptions are urgent and users may not want to or have time to leave cues.

Upon resumption computers can help people find information or cue memories about the suspended activity. Prior work in this area has focused on the three themes of personal information management, activity based computing, and rich activity histories. Personal information management systems such as Stuff I've Seen and Haystack sought to collect and organize users information including email, contacts, calendar events, and documents so they could quickly navigate a web of related information to find information they were looking for (Karger, Bakshi, Huynh, Quan, Sinha, 2005; Dumais et al., 2003). Such systems are particularly useful when trying to recall connected information, such as who was on a particular conference call last week. However, they are focused on restoring the information that computers have access to rather than the

ephemeral mental context that workers had at the time of interruption.

Motivated by the recurring finding that people think of their computer use in terms of overarching activities rather than individual applications or tasks (Suchman, 1986; González & Mark 2004), many activity based computing systems have tried to make it easier for people to manage information by activity and thus switch between activities more quickly. The seminal Rooms system supported users in arranging documents in separate desktop "rooms" associated with specific tasks or activities (Henderson & Card, 1987). Since then a variety of projects such as Scalabe Fabric (Robertson et al., 2004), Taskposé (Bernstein, Shrager, & Winograd 2008), Giornatta (Voida, Mynatt, & Edwards 2008), Task Tracer (Dragunov et al., 2005), CAAD (Rattenbury & Canny, 2007), UMEA (Kaptelinin, 2003) and Bardram's Activity Based Computing environments (Bardram, Bunde-Pedersen, & Soegaard, 2006; Houben et al., 2013; Jeuris, Houben, & Bardram, 2014) have provided ways for users to manually or automatically cluster their windows, documents, and communication streams by activity. A recurring finding of this work has been that it is difficult to accurately and automatically cluster windows by activity, or to motivate people to spend much time organizing their activities manually.

Researchers have also worked to make it easier for users to re-find relevant information about suspended activities by creating rich activity histories. Projects in this direction include automatically generated laboratory notebooks such as Prism and Burrito (Tabard, Mackay, & Eastmond, 2008; Guo & Seltzer, 2012) and extensive application histories such as Chronicle or Youpivot (Grossman, Matejka, Fitzmaurice, 2010; Hailpern et al., 2010). These systems are built around finding information (e.g., What was that paper I was reading while listening to AC/DC?) or sequences of actions (e.g., What sequence of commands produced this graphic effect?) but not helping people recall what they were thinking.

2.6. Visual Histories

One promising approach to restoring mental context lies in cuing users own memories with images from their past. Humans have remarkably discriminative visual memory. In one study, people shown a series of 2560 images for 10 seconds each were able to correctly select 90% of the time which of two images they had seen before when tested two days after initial presentation (Standing, Conezio, & Harber 1970). A similar study found that subjects were able to correctly identify the image they had seen before 87% of the time, even when the only difference between images was the pose of an object, such as the arrangement of beads on an abacus (Brady, Konkle, Alvarez, & Oliva 2008). While the images in these studies were all of generic objects, this evidence suggests that people might be able to discriminate between relatively similar images of their past activities.

This potential of images to cue episodic memory is also supported by more applied research. Lamming and Newmann argued that researchers could benefit from considering the nature of autobiographical memory when designing tools to help people recall past activities (Lamming & Newmann, 1992). In particular, they noted that autobiographical memory, or memory about one's past, is best cued by stimuli that were present at the time of memory formation. Lamming and Newmann showed that people remember more

about their days while reviewing time-lapse videos than when relying on abstract textual summaries or rote memory. More recent studies have corroborated this finding with videos of users in their office (Czerwinski & Horvitz, 2002) and images taken with wearable SenseCams (Sellen et al. 2007). Work with programmers has similarly revealed that showing snippets of recently edited code leads to faster resumption of programming tasks (Parnin & Deline, 2010) and more accurate identification of past activities (Safer & Murphy, 2007) than reviewing lists of recently edited functions or files.

While these studies demonstrate that people can recognize their prior activities in images of those activities, there is less evidence that they can go a step further to remember what they were thinking at the time. Case studies have shown that individuals with memory impairments are able to remember more details about past events when viewing images of those events, as compared to relying on unaided memory or hand-written diaries (Hodges et al., 2006, Lee & Dey 2008). However these studies used images of distinct events (such as visiting a museum or restaurant) that had been taken with a wearable camera and curated by a caregiver. Moreover, the details measured included things like where participants were and whom they met, which may be very evident in the image. It remains to be seen if images of their past can help people recall what they were thinking, particularly if the activity is computer-mediated and the information is not evident in the images themselves. The first study, described next, explores this possibility.

The following studies build on prior work demonstrating that 1) visual histories of past work can cue users' memory for contextual details that are not present in the histories themselves, 2) in some cases lightweight, easily recorded and reviewed visual histories are sufficient to cue detailed mental context, and 3) that the cues, artifacts, and strategies knowledge workers currently use to restore mental context vary across activities and disciplines.

3. RECONSTRUCTING CONTEXT WITH DESKTOP THUMBNAILS

People remember more of their prior activities when they see images or videos of those activities than when they rely on textual summaries or unaided memory (Lamming & Newmann, 1992; Czerwinski & Horvitz, 2002; Sellen et al. 2007). This recurring finding suggests that visual histories can help people recall suspended activities. It remains to be seen if they might also help people reconstruct the rich mental context associated with these activities.

Even if visual histories can be designed to reliably cue rich mental context, and thus aid resuming complex activities, there are at least two practical barriers to using them for that purpose. First, recording visual histories can be burdensome. The visual cues used in past studies have primarily been images of participant's physical environments. These images include details about the participant's location, posture, and work materials, which may aid reconstruction of detailed context. However, obtaining them requires using wearable or wall-mounted cameras, which, while feasible for short-term studies, may be seen as invasive or cumbersome in the long-term. Second, reviewing visual

histories can be burdensome. Large collections of images or video of past work can be computationally expensive to record and tedious to review. Some previous studies used caregivers to curate a set of meaningful images for memory impaired loved ones (Hodges et al., 2006; Lee & Dey 2008). Lightweight methods of capturing, summarizing, and displaying activity are needed if visual histories are to be sustainable for representing everyday activity at scale.

The following study tests the context-cuing capability of *desktop thumbnails*, small images and short videos of prior computer activity. Thumbnail images are widely used to represent and aid re-finding previously visited websites. Their small size means they are easy to record, store, and arrange in scannable layouts. To see if small, lightweight desktop thumbnails might help viewers restore mental context, we had six people record their everyday work activity for two weeks and describe their prior work while viewing images and videos representing that activity.

3.1. Methods

Participants. We recruited research colleagues for this initial study due to the potentially invasive screen recording involved. Six graduate students (aged 23-29, 4 females) agreed to participate.

Materials. Each participant installed custom screen recording software on a single work computer. This software took an image whenever they clicked, typed, or paused for more than a few seconds. This method of time-lapse recording enabled us to capture a fairly complete visual history of a participant's work without significantly slowing down their computer. Recorded images were stored on a USB drive. The computers participants used for recording ranged from 13" laptops with a screen resolution of 800 x 1280 pixels to 27" desktops with a resolution of 1440 x 2560.

We used participants' time-lapse recordings to create 8 different types of thumbnail, varying thumbnail design along three dimensions in a 2x2x2 design. These dimensions were: animation, focus, and familiarity. Together these design variations tested how lightweight recording and representation of computer activity could be and still cue rich mental context.

> *1. Animation: static screenshot or video.* By displaying activity over time, videos of computer activity provide more information than individual screenshots, potentially helping users distinguish similar activities and retrieve more specific memories. However, making sense of videos requires intense focus (Czerwinski & Horvitz, 2002) and may cause viewers to neglect mental context, leading them to only recall aspects of the activity that are visible in the video. Screenshots may provide less detail, but encourage viewers to elaborate from memory. We tested this variation to see if static images could cue rich mental context, or if videos (which take more computation to record and attention to review) are needed.

> *2. Focus: full screen or cropped.* In pilot tests we noticed a tradeoff between full screen thumbnails and thumbnails focusing on a small area of the screen. Images of a

full desktop reveal window arrangements, which can help viewers quickly identify an activity. However it can be difficult to distinguish details if these images are small, such as a section header or the first few lines of a function. Activities such as writing and programming may be better represented by thumbnails showing only a small portion of the screen. These regions of interest might be identified by detecting the user's cursor location, gaze, or inspecting the application's visual structure (Figure 2). We tested this variation to see if images showing small portions of a user's screen could cue rich mental context, or if the entire screen needs to be represented.

3. Familiarity: thumbnail reviewed or not. Reviewing memories makes them easier to recall later, but can also modify their contents (Hupbach, Hardt, Gomez & Nadel, 2008). Asking people to briefly review thumbnails of past activities before they need to resume those activities may make details about those activities easier to recall when they review that thumbnail again, but may also restrict the number of details they can remember. We tested thumbnails representing moments participants had already seen an image of and reflected on, and moments they had not seen an image of or reflected on, to see if users need to be actively involved in recording and reflecting on their activity or if more passive methods of recording are sufficient to produce meaningful contextual cues.

(Figure 2 about here)

Procedure. Participants were asked to record their Monday-Friday work activity for two weeks, pausing recording whenever they used video chat or were performing a private activity such as online banking.

We used periodic experience sampling to sample participants' activities and end-of-day debriefing to prompt reflection and collect baseline measures of their mental context for these activities. Every 30 minutes, the recording software prompted participants to answer the question "*What are you doing?*" in a simple text box (Figure 3). Participants could ignore these prompts by immediately hitting an *Enter* or *ESC* key, but were encouraged to respond to as many as possible. At the end of each day, participants used a debriefing module to provide additional detail on up to five moments sampled earlier that day (Figure 3). While viewing their written description of the moment and an accompanying screenshot, participants were asked to answer the following questions:

1. What do you remember about this moment? (free text or recorded audio response)

2. How do you know these details? (I remember this moment exactly, I know from experience, or I'm guessing)

3. How well does this image represent what you were doing? (Very well, Somewhat, or Not at all)

This two-part sampling was designed specifically to collect data about participants' mental context without disrupting them in the moment and keeping interruptions as minimal as possible. Prior work by Brandt, Weiss, and Klemmer (2007) has shown that

asking people to write short descriptions in the moment and fill in details later reduces the interruptiveness of experience sampling but maintains the quality of responses. For example, while going through papers after lunch one day, P1 received a prompt asking him "What are you doing?" to which he responded "collecting literature for dissertation". During a debriefing session later that day, while viewing his description and a screenshot of this moment, (Figure 3) he elaborated "I was trying to find papers that had specifically cited a book I'm using for my dissertation. I was also kind of annoyed because a friend was chatting with me online. This particular paper did not end up looking very helpful." Note that there is no evidence in the screenshot that the participant had been chatting online with a friend, but he still remembered doing so and how he felt about the interaction.

(Figure 3 about here)

At the end of each week of recording, we interviewed participants about their experience using the recording software and had them review up to 40 desktop thumbnails (up to 80 across the two weeks) representing moments from earlier that week. We tested four types of screenshot thumbnail during the first week of the study, i.e., full-screen and cropped images representing both moments participants had reflected on during debriefing sessions and moments they had not (no animation condition). During the session, participants were first presented with a blank screen. Similar to Kaasten, Greenberg, and Edward's test of website thumbnail size (2002), when the participant pressed a start button, a 20px high thumbnail would appear. (The width of the thumbnail depended on the recording computer's screen ratio). Every two seconds the thumbnail grew 20px taller and proportionately wider. Participants were instructed to press a stop button once they recognized the represented activity (e.g., editing a particular paragraph) rather than just the overarching project (e.g., writing a conference paper). Once participants recognized the activity, the thumbnail stopped growing. They then answered the same three questions posed during their end of day debrief sessions (i.e., What do you remember…?, How do you know…?, How well…?). Full screen thumbnails always showed the participant's entire screen, though at low resolution at first, whereas cropped thumbnails showed an expanding area around the user's cursor location in full resolution (i.e., if the recorded screen had a 10:16 ratio, the cropped thumbnail would first show the 20x32px area surrounding the cursor, then the 40x64px area, and so on). After completing each response participants would press the start button again and be presented with a new thumbnail.

We tested the four types of video thumbnails during the second week's review session (i.e., full-screen and cropped videos representing both moments participants had reflected on during debriefing sessions and moments they had not). Each video showed a time-lapse of five minutes of computer activity played at 5x speed. The goal with these videos was not to determine how large the videos had to be, but how long they had to play before participants recognized their activity. Prior work on website thumbnails has recommended setting the default thumbnail size to the height at which 80% of websites can be recognized. During the first week of our study, our first four participants recognized 80% of their activities when cropped thumbnails were 520px tall and full-screen thumbnails were 320px tall. Favoring recognizability, we used the 520px height

for the height of both sets of video thumbnail. While relatively large for a thumbnail, this was still significantly smaller than our participants' average screen height of about 1200px. Full-screen videos showed a 520px tall time-lapse of the participant's full screen, while cropped videos showed a time-lapse of the 520px tall area surrounding the participant's cursor at full resolution. Participants pressed a button to start each video, pressed a second button which stopped the video once they recognized the represented activity, and then filled out the debrief questions before moving on to the next video thumbnail.

Analysis. We automatically collected the thumbnail size and video duration at which participants stated that they recognized their activity. A large thumbnail size would mean that images have to be large before people can use them to reconstruct context, reducing their ability to be composed into collages or scannable layouts summarizing a day's activity. Similarly, long video durations would suggest people need more time to review videos of past activity before being able to reconstruct context, limiting their ability to quickly cue memory. We also tracked participants' ratings of their memory strength (I remember this moment exactly, I know from experience, or I'm guessing) and how well the thumbnail represented their activity (Very well, Somewhat, or Not at all). We wanted to see which thumbnail variations were rated as very representative of prior activity and as cuing strong memories.

To get a more objective measure of the accuracy and detail of participants' memories, we hand-coded each verbal and written description of mental context cued by each thumbnail. We coded memory accuracy on a scale of how well the activity described in the end-of-week thumbnail response matched the activity described in the end-of-day debrief. The levels included "mostly matches" "partial match", "no match". Accuracy could only be coded for familiar thumbnails since unfamiliar thumbnails had no baseline from end-of-day debriefs for comparison. We also coded the number and type of contextual details included in each response. This included the type of detail (who, what, where, when, why, or feeling information), the level of activity described (project, activity, or task) and whether the described events occurred before, during, or after the sampled moment. An example of a project description would be "working on my presentation", an activity might be "editing this discussion slide", and a task might be "copy and pasting references".

Two researchers coded the end-of-day debrief responses from the first week of the study and iterated the coding rubric until they achieved a Cohen's Kappa of >0.60 for each item. They then separately coded participants' responses to the end-of-week thumbnails. Across the two weeks of the study, participants responded to an average of 36 thumbnails or about five thumbnails representing each of the eight thumbnail design conditions; though there was significant variation in the number of times each participant saw each type of thumbnail. To account for this repetition and variation in our analysis, we fit a mixed-effects linear model to each dependent measure, using animation, focus, and familiarity as our fixed effect variables and participant number as a random effect variable. The significance of each variable's impact on the final measure was determined

with a Likelihood Ratio Test.[1]

3.2. Results

Across the two weeks of the study, participants recorded 210 hours of their everyday work activity, averaging 17.5 hours per participant per week. Participants recorded a range of activities including writing a thesis, drafting a conference presentation, labeling brain scans, analyzing data, and shopping for a car. While recording, participants saw 405 experience sample prompts and responded to 400 of them in-the-moment with a short description of what they were doing. They provided additional detail on 188 of these moments with text or audio descriptions during end-of-day debrief sessions. During the end-of-week lab sessions we tested a total of 145 image thumbnails, and 70 video thumbnails with participants (two participants had too few experience samples during the second week of the study to participate in the video thumbnail testing). Half of these thumbnails represented moments participants had reflected on during their end-of-day debrief sessions (i.e., familiar thumbnails), and half represented moments they had not previously reflected on (i.e., unfamiliar thumbnails).

Cuing Mental Context with Desktop Thumbnails. The first aim of this study was to see if participants could reconstruct rich mental context while viewing visual histories of their past computer mediated activity. We were not sure which type of visual history would be most effective, so we tested eight variations of thumbnails. Participants were remarkably consistent in the number and type of contextual details they were able to reconstruct while viewing each type of thumbnail.

Across the eight variations, participants' descriptions of past activities averaged 55.2 words ($\sigma=30.1$) in length and contained 4.47 ($\sigma=1.89$) contextual details (i.e., who, what, where, when, why, feeling information). Most responses described what participants were doing at the time (99%), though a large proportion of responses also described why an activity was taking place (47%), who was involved (40%) and when it occurred (38%) (Figure 4). A smaller, but non-trivial, number of responses included where participants were (9%) and how they were feeling (12%). Participants were consistent across conditions in how often they discussed past events (29% of responses) or future events (24%) with nearly half (47%) of all responses including a reference to either the past or future.

Participants marked 75% of thumbnails as representing the moment "very well", 23% as "somewhat" representing that moment, and only 2% as "not at all". Participants also

---

[1] Running a repeated measures ANOVA would require excluding much of the data to balance the number of times each participant responded to each type of thumbnail. Mixed effect models allow for repetition and variation in the number of times participants are exposed to each condition while accounting for both fixed effects caused by manipulating independent variables and random effects caused by individual differences. The likelihood ratio test checks the impact of each independent variable by removing it from the model and using a one-way ANOVA to test for differences between the values predicted by full and reduced models. (Seltman, 2012)

stated that for 50% of the thumbnails they reconstructed contextual details "from experience", while 43% of the time they "remembered exactly", and 7% they "guessed". In our own ratings of memory accuracy, we found 60% of memories cued by desktop thumbnails "mostly matched" participant's day-of memories, 35% "partially matched", and only 5% did not match at all.

(Figure 4 about here)

The second aim of this study was explore whether lightweight, unobtrusively recorded thumbnails can cue recollection of rich mental context, or if larger, more detailed, and more actively recorded representations of prior activity are needed. The following sections describe the significant differences we observed in mental context reconstructed by various types of thumbnail.

Effectiveness of Images vs. Video. Participants tended to rate memories cued by video thumbnails as stronger than those cued by image thumbnails. They were 1.81 times more likely to give memories cued by videos a stronger rating than those cued by static images, though this effect was not robust (95% CI [0.57, 5.78]). Despite participants tending to believe memories cued by video thumbnails were stronger, their descriptions of these moments were much less likely to match the descriptions they shared during end-of-day debriefs. Their responses to videos were only 0.17 times as likely (95% CI [0.04, 0.67]) as those prompted by images to achieve higher accuracy ratings. Responses to video thumbnails were also substantially shorter, averaging 17.3 fewer words (95% CI [-3.6, -30.9]).

Effectiveness of Cropped vs. Full Thumbnails. Participants were 4.09 times more likely to rate full screen thumbnails as representative of their activity than cropped thumbnails (95% CI [1.32, 12.67]). This difference was reflected in the fact that participants needed to see larger versions of the cropped thumbnails before they could recognize their activity. On average, participants needed cropped thumbnails to be 142.1 pixels larger than full-screen thumbnails before they could recognize their activity (95% CI [68.0, 216.2]). The average cropped thumbnail had to be 370px tall (σ=210px) before participants recognized their activity as opposed to only 240px tall (σ=154px) for full-screen thumbnails. This difference between cropped and full-screen thumbnails was not apparent in the duration of videos needed to recognize an activity as the 1.3 fewer seconds needed to recognize an activity in a cropped video was not robust (95% CI [-3.9, 1.9]). With full-screen videos participants recognized their activity after an average of 8.0 seconds (σ=4.1s) and after 8.1 seconds (σ=5.3s) with cropped videos.

Effectiveness of Familiar vs. Unfamiliar Thumbnails. There was no significant difference in the length, content, or accuracy of mental context cued by thumbnails that had been previously reviewed during end-of-day debriefing sessions and those that had not been reviewed. Nor was there any measurable difference in the size of images or duration of videos needed to recognize the activity.

There was only one significant interaction effect in which responses to cropped video thumbnails averaged 1.4 fewer details than those cued by other types of thumbnails (95%

CI [-2.8, 0.0]).

3.3. Discussion

Can visual histories cue reconstruction of rich mental context for past computer-mediated activity? With the exception of cropped videos, participants were remarkably consistent across conditions in the number and type of details they shared while viewing desktop thumbnails. Many responses included details about not only what participants were doing, but also why and what they were thinking at the time. A typical response was:

> "So here, this is another one from one of my intense lit searches last week. I can tell that I'm looking in Google Scholar for a paper from 1909, which I didn't have very much hope of finding. But it was probably one of the earlier ones that I thought would just be listing lots of phenomena that I thought would be worth talking about, because that's what these older papers usually are. But I highly doubt that I ended up finding it because it was really old." – *P1's response to a 160px high cropped image thumbnail (Figure 5).*

(Figure 5 about here)

In this example, the thumbnail itself is fairly sparse, a few dozen words of text with simple formatting (Figure 5). Yet the participant was able to recognize that the formatting matched a typical Google Scholar page, and deduce that the words "Jespersen" and "1909" had been bolded because he had searched for these terms. From there, he was able to reconstruct 1) his activity ("intense lit search"), 2) his task ("looking in Google Scholar for a paper from 1909"), 3) why he was doing the activity ("it was probably one of the earlier ones that I thought would just be listing lots of phenomena that I thought would be worth talking about"), and 4) make an educated guess as to whether he successfully completed it ("I highly doubt that I ended up finding it").

This result demonstrates that, given the right visual cues, people can reconstruct remarkably detailed mental context. Moreover these memories were reasonably accurate as 95% of them "mostly" or "partially" matched participants' day-of memories.

How might we design lightweight but effective visual histories? Of the variations we tested, the simpler thumbnails (i.e., static full-screen images) were as or more effective at cuing detailed and accurate mental context than more complex thumbnails involving video recording or end-of-day reflection. With random selection of activities and some rather small and sparse thumbnails, participants were still able to consistently reconstruct non-trivial context about what they were doing, thinking, and feeling at the time. A paradox emerged from our results regarding video thumbnails. Participants were more confident that they remembered exact moments when they viewed videos of those moments, but these memories were less similar to their day-of responses than memories cued by static images. One possible cause of this mismatch between perceived and actual accuracy was that in several cases participants simply described what was happening in the video and did not elaborate with other contextual details. For example, compare the

following end-of-day debrief and end-of-week response cued by a video thumbnail. The debrief focuses on motivating events and talks about the activity abstractly ("analyzed the data"), whereas the video response is much more specific ("making a pivot table"), but fails to mention the broader context or motivation.

> *End-of-Day Debrief:* "So I released a batch of twenty participants through Mechanical Turk this morning at 9am and for some reason the traffic was very slow today. So until 2pm I think I had only gotten about ten people. But then I didn't want to wait until I get all twenty so I went ahead and analyzed the data just to see the pattern and saw that it wasn't in the direction that I wanted."

> *Response to Video Thumbnail*: "So I was making a pivot table and I was trying to see if I saw any pattern from my data, but I'm not sure if it was before I got all my data or after."

Cropping images to show a region of interest in more detail did not enable us to cue richer mental context with smaller thumbnails, at least not when naïvely using the user's cursor location to position the crop. In many of our cropped thumbnails, the user's cursor was over a blank portion of their screen, and the thumbnail had to grow rather large before it showed anything meaningful. This may explain the 141 pixel difference in the height of full-screen and cropped thumbnails needed to recognize an activity. It may also explain why participants provided fewer details in response to cropped video thumbnails as these videos may have included fewer meaningful cues. As Sellen et al's work (2007) with SenseCams demonstrated, in many cases, when viewing an image from their past people do not so much "remember" the moment as "know" it by combining cues from the images with general knowledge about their own habits. Indeed, 50% of the time our participants said they knew context from experience rather than remembering an exact moment. Our naïve cropping seemed more often to remove reconstructive cues than to focus on them, as was reflected in participants being less likely to say that cropped thumbnails represented the activity well.

However, a tradeoff remains between showing full desktop arrangements and the details of a particular activity shown in cropped thumbnails. After responding to each thumbnail we gave participants the opportunity to describe what part of the thumbnail sparked their memory. These self-declared cues included high-level features that are easily detected in small full-screen thumbnails, such the general look of an application or window arrangements, but also low-level features that can be hard to see in small full-screen thumbnails such as text, folder names, and images. More research is needed to determine the tradeoff between high and low level visual features and how meaningful low-level features may be automatically identified in an image.

Combined, these results suggest that static images of full-screen desktop arrangements are better cues than cropped images or videos, leading to more detailed and accurate recollection of mental context. This result is particularly encouraging when compared to past website thumbnail studies. Kaasten, Greenberg, and Edwards (2002) found that people recognized 80% of websites by the time the thumbnail was 208px tall. We found that participants similarly recognized 80% their activities *and* remembered non-trivial

mental context about them by the time full screen image thumbnails were 320px tall. Taking into account that Kassten et al. used a 868px tall monitor whereas our average participant's monitor was about 1200px tall, Kaasten's website thumbnails had to be 24% of the screen height whereas the desktop thumbnails we tested had to be only slightly larger at 26% of screen height to consistently cue mental context. As a final note, neither the size of thumbnails nor the detail of memories was significantly impacted by whether or not the participant had debriefed on the moment before. That is, the end-of-day debriefing sessions did not seem to make participants notice their activity with larger or smaller thumbnails, or to remember mental context that was any more or less detailed than for moments they had not reviewed before. This finding agrees with Sellen et al.'s (2007) result that randomly taken SenseCam images were as effective at cuing memory as images taken explicitly by users. It may not be necessary then for users of visual histories to engage in reflection or memory consolidation for those histories to be useful for reconstructing mental context at a later time.

Perceived burden of recording visual histories. Our participants were highly engaged in the study, recording 17.5 hours of computer activity per week and responding to nearly all experience samples. This may be due to the fact that, as research colleagues, they wanted to provide "good data". Many stated as much during interviews. However, they also revealed that generally the recording, experience sampling, and debriefing felt non-invasive and even prompted reflection.

Three of the six participants said that the recording did not negatively affect their work but rather helped them avoid distracting websites such as Facebook, or to be more intentional about switching between working and relaxation. Three participants kept the recording on for the entire duration of study. Four stated that the experience sampling was not burdensome, as they may have expected. On the contrary, two participants mentioned that the periodic experience sampling helped them keep track of time and consider if they should be spending so much of it on a particular task. Four mentioned that the end of day debriefing helped them reflect on their day or think about what they needed to do the next day.

These responses may be biased and need to be validated with longer-term deployments. Nonetheless, they suggest that lightweight ways of capturing and reflecting on everyday activity, both in the moment and at the end of a day, help people think critically about their work.

4. OBSERVING HOW WRITERS, PROGRAMERS, AND GRAPHIC DESIGNERS RECONSTRUCT MENTAL CONTEXT

The first study demonstrated that lightweight visual histories, such as small thumbnail images of prior desktop arrangements, support both the recognition of prior computer-mediated activities and reconstruction of associated mental context. Participants were able to notice small features in these images, such as unique window arrangements or distinct text, which led them to reconstruct detailed memories about what they were thinking and doing at the time. This first study hinted to the importance of visual cues in triggering recognition, raising the question of how such cues could be recreated from

computer records. In addition, it raises more fundamental questions about the types of mental context that knowledge workers need to recover when resuming a suspended activity, and how this varies by activity.

Prior research on how knowledge workers restore mental context has largely focused on programmers. Programming often involves manipulating large collections of interconnected text files to achieve pre-defined goals. As a result, discussions of mental context have focused on how people remember suspended goals or connections between files and blocks of code (Parnin & Deline 2010). Resumption aids such as the Mylyn extension for the Eclipse IDE, which filters the items shown in Eclipse's file tree based on a user's self declared task, are a result of this line of work (Kersten, Elves, & Murphy 2006). These aids are useful for programming, but not necessarily for other types of knowledge work.

Moreover, programmers' mental context seems to be more complex than a list of goals and files. In our first study participants associated a wide range of information with programming tasks, including goals and files, but also whom they were working with and how they were feeling. Past research also has focused on explicit actions such as leaving inline comments or intentional bugs to aid context reinstatement. However, it has largely overlooked the unintentional and opportunistic ways in which knowledge workers store and restore mental context.

Motivated by results from our first study, we wanted to explore how people reconstruct mental context across a range of domains and activities. Specifically we sought to understand 1) what kinds of *information* knowledge workers seek when resuming various activities, 2) how this information is *represented* in memory or artifacts, 3) what *strategies* knowledge workers use to reconstruct context, and 4) how the information, representation, and strategies *depend on the activity*. Answers to these questions have the potential to help identify meaningful moments in a stream of recorded activity and represent those moments with meaningful memory cues. We observed fifty resumption episodes, gathered by screen recording and interviewing web developers, scientific writers, and graphic designers as they went about their everyday work. We find participants performing each type of activity engage in situated sensemaking, constructing stories to explain implicit cues and reconstruct rich mental context.

4.1 Methods

Participants. We observed ten people (aged 19-30, 8 female) as they resumed graphic design (2 participants), web development (7 participants), and scientific writing activities (3 participants). Note that some participants pursued more than one type of activity. All participants were university employees or students recruited from a project-based web design course, a dissertation writing workshop, or a campus user experience group.

Materials and Procedure. We asked participants to record their Monday-Friday work activity for two weeks using custom screen recoding software that took a screenshot whenever they clicked, typed, or paused for more than a few seconds. Most participants installed the recording software on their laptop computers, enabling capture of

resumptions occurring throughout the workday, but also in the early morning or late evening in cafes and personal residences. The screenshot sampling captured detailed time-lapses of each participant's computer screens without impacting use of computationally intensive programs such as Photoshop. We asked participants to leave short textual notes to mark moments when they were resuming a long-term activity using a note-taking feature built into the recording application. We additionally asked participants to pause recording whenever they used video chat or were performing a personal or private activity, such as online banking.

We interviewed participants for about an hour at the end of each week of recording. During each interview we had participants find and review portions of their screen recording when they were resuming a long-term activity and describe what they were thinking and doing at the time. Previous work has found such auto-confrontation to be an effective means of obtaining details about a recorded moment without requiring much input from the participant at the time of recording (Cangiano & Hollan, 2009). Participants selected specific moments to review using a visualization that showed their screen recording, their general level of computer activity (i.e., clicks per minute), and their textual notes (Figure 6).

(Figure 6 about here)

Data Analysis. We transcribed participants' interview descriptions of each resumption episode and also analyzed the 15 minutes of screen recording following each resumption. We selected fifteen minutes based on prior estimates of how long it takes to reconstruct mental context (van Solingen, Berghout & van Latum, 1998). Three researchers reviewed this corpus and iteratively coded the transcripts and screen recordings for the types of information sought, how they were represented, and strategies for interacting with representations. In the results that follow, we first describe the types of resumptions observed and then the information, artifacts, and strategies involved in resuming mental context. The results demonstrate that mental context is routinely more complex than a list of goals, and it is often restored with the help of implicit cues that are the natural result of prior work, not intentional explicit cues such as notes.

4.2. Results

Types of Resumption

Across the two weeks of the study participants recorded 229 hours of computer activity, or about 11.5 hours per participant per week. Across the ten participants we collected fifty resumption episodes: ten of graphic design activities, thirteen of scientific writing activities, and twenty-seven of web development activities. These interrupted activities included maintaining a university website, developing web applications, writing a masters thesis, creating several seminar posters, and designing a conference program booklet. Since participants may not have recorded all of their computer activity during the study and we did not record their work activity away from their computers, we cannot know with certainty what participants were doing immediately before resuming their activity in all cases, in only thirteen of the fifty resumption episodes did participants

record their activity immediately before the resumption. This may be because participants were likely to be away from their computer just before resuming a substantial activity, or because they chose to only turn on recording for certain activities. Based on analysis of the screen recordings across the fifty resumptions, participants were resuming their activity after an average of 42 hours away (σ=48 hours) though there was significant variation. In only twenty of the fifty resumption episodes was the activity being resumed the same as the activity that had last been recorded.

Information

Characterizing the restoration of mental context as a process of reconstruction raises the question: What are the basic building blocks of that construction process? While the specifics likely vary by activity, (e.g., a designer may need to recall if a document has an RGB or CMYK color scheme whereas a programmer may need to remember whether a function uses recursion) a more abstract characterization may reveal where activities overlap. As in prior work, we observed participants recalling suspended goals, plans, actions, and the artifacts needed to enact them. However, this was only part of what participants described as being necessary to resume interrupted activities. In addition, participants reconstructed information about collaborators, the history of the activity, and connections between collaborators, goals, plans, actions, and artifacts. This paints a much richer picture of mental context than prior work. We detail the findings below.

Goals. How people remember *a suspended goal* has been a key focus of the psychological literature on interruption and activity resumption (e.g., Altmann & Trafton, 2002). However, in their descriptions of resumption episodes our participants rarely recalled a single goal in isolation. Instead, we observed them recalling networks of goals. Some networks were hierarchical as when P3, a graphic designer, described her suspended activity as achieving both the high-level goal of "finish the seminar poster" and the low-level goal of "find the speaker photo". Other goal networks were more complex in that achieving one goal helped achieve several others. For example, P2, a web developer, had the goal of "determine what information goes on the My Events page" which supported not only the goal of "develop the My Events page" but also "implement a database".

Participants reconstructed several types of goals including objective goals, process goals, and constraints. Objective goals stated what they wanted to achieve (e.g., "fix the rendering bug"). Process goals involved an intermediate step in a process that needed to be taken rather than a final state to achieve (e.g., "figure out how the rendering currently works"). Finally, participants recalled constraints that had to be met, but did not directly contribute to achieving the objective goal (e.g., "turn the code in by midnight").

Plans and actions. Plans were also frequently reconstructed, but often as a general approach to achieving a goal rather than an exact sequence of actions. P1, a web developer mentioned "We were thinking about using Google Drive as a backend for the notes" even though her team had not outlined the exact steps for doing so, or even determined whether the plan was feasible. In addition to the suspended goal and plan, participants reconstructed their last suspended action as a low-level description of what

they needed to do next. When resuming one programming activity, P2 first recalled that he needed to download his collaborator's codebase from Github (the immediate next action) and then remembered that this was so he could compare it with his own version of the code to determine why a page rendering bug appeared in his code but not that of his collaborator.

Artifacts. Past work on activity based computing has focused on helping people organize and restore digital artifacts related to a particular activity. Our participants often restored not just the artifact itself, but also mentally reconstructed information about its existence, location, and contents. Recovering artifacts and reconstructing information about them from memory were often simultaneous and entangled processes. For example, P10, a writer, first recalled the location of a set of meeting notes on her computer and, upon viewing a graphical preview of those notes in her OSX Spotlight search, realized she had not actually used them during the meeting, so they would not contain any information about what was actually discussed.

Collaborators. Past descriptions of mental context have largely assumed a single actor pursuing an activity in isolation. However, knowledge work is routinely collaborative. This raises at least two challenges in regards to reconstructing mental context. First, information about collaborators as well as their activities needs to be reconstructed. Specifically, we observed participants recalling a collaborator's skills, knowledge, access to information, and relationship with other collaborators. For example, P2, a web designer, needed to consider his collaborator's familiarity with version control when deciding how to assign and coordinate his team's development activities. Collaboration also raises an additional challenge in that the state of the activity may have changed significantly while the worker was away. Goals may have been accomplished, plans changed, actions taken, and artifacts modified, created, or destroyed.

Attributes and connections. As implied above, mental context consists not only of specific goals, plans, actions, artifacts and collaborators, but also information about these entities. Examples of these attributes include the priority of a goal, estimated cost of a plan or action, location of an artifact, and skills of a collaborator. Moreover, we rarely observed participants restoring a particular aspect of context in isolation. Instead they used the attributes of one aspect of context to reconstruct connections to other aspects. For example, P1 noticed her collaborator was modifying all of their HTML code to align with the Handlebars templating framework, realized this would involve making changes to most of the codebase, determined that she was not comfortable enough with version control to merge her code with those changes, and decided to wait to download the latest version of the code until after her collaborator had made his changes. Simply recognizing which action her collaborator was taking was not enough information to decide what she should do. She also had to consider the impact of those actions on shared artifacts and their relation to her own skill set.

History of activity. A final type of context we observed was recalling the history of an activity. Of particular interest was recalling past states of artifacts, comparing their current state to the last seen state, and comparing a current state to a projected future state that would meet a particular goal. P2, a web developer, needed to recall prior names his

team had given their project so he could find and replace all instances of those names in their code. Upon viewing the current version of a seminar poster, P3, a graphic designer, realized she only had to update the talk description before the poster was "finished".

Representation

Knowledge workers have numerous means of preserving mental context including writing notes, placing cues in the environment, or simply remembering specific information. The representations participants used varied considerably, both across participants and activities. Targeted studies, such as the programmer-centric studies discussed above, are needed to determine the exact mechanisms knowledge workers use to represent context for individual activities. Here we focus on the distinction of how explicitly information was represented and highlight the value of implicit cues.

Explicit representation. Some mental context was explicitly represented in that it could be read directly from a physical or digital artifact. For example, goals were frequently written down in todo lists. All ten of our participants used todo list to keep track of activities and many kept separate lists of goals for each activity. One example is P1's list of todos for a class project shown in Figure 7, which she shared online with teammates. P4, a writer, used a physical notebook to keep track of not only pertinent quotes from articles she had read, but also her interpretation of those quotes so she did not have to remember why she thought they should be included in her thesis.

(Figure 7 about here)

Explicit context was not always stored in separate organizing artifacts but was often placed in key artifacts by individuals or their collaborators. P6 noticed an inline comment in a program file written by a collaborator that let him know that a bit of code still had bugs. P10 left notes for herself in a LaTeX document using a built in commenting feature.

Explicit context was also present in communications between collaborators. P1 used instant messaging to ask her teammates if one of the items on their shared todo list was completed after seeing that the item's status was "Working" which could mean either that the code was working, or was being worked on. Rather than search through online documentation, P3 walked down the hall to ask a coworker about the resolution of the TV that would show a digital sign she was creating. All participants frequently referenced emails for information about ongoing activities.

Explicit context played an important role in facilitating collaboration. Information that had to be shared with teammates including the location of shared artifacts and the status of completing goals, plans, and actions, was typically represented or communicated explicitly.

Implicit Representation. Despite the prevalence of explicit representation, a substantial portion of mental context was reconstructed from implicit cues. Implicit cues are ambiguous and often a byproduct of past work. Information about actions,

particularly which action to do next, was often represented implicitly. For example, Figure 8 shows part of a hand-in sheet that P1 and her teammates were filling out for a class assignment. When P1 first looked at this document, she was still figuring out what to do next. She was able to use two implicit cues to determine that her team needed to come up with a "Task-to-Try" to demo in class the next day. While all the other headings in the document had information under them, the Task-to-Try section was blank. She also observed that the cursor of one of her teammates was focusing on the "Task-to-Try" part of the document. This sharing of attention was not done intentionally by her teammate, but was a byproduct of using Google Docs.

(Figure 8 about here)

Implicit cues showed up in writing and graphic design as well. While working on a qualifying paper, P10 was able to tell where she left off writing last time by finding the point where free-text gave way to outline. While designing a pamphlet for an upcoming conference, P7 was able to tell where she needed to start editing the conference flier by seeing which sections of the flyer's schedule page were still blank. P3 intentionally left herself an implicit cue by writing "Something about teaching" in place of the title on a seminar poster. This acted both as an implicit cue that she needed to update the title, since looking at the title she would know that it did not conform with typical title language, and as an explicit cue that the title would be about teaching.

Participants also identified implicit cues by comparing artifacts with memory. Several participants working on web development (P1, P2, P4) stated that they were so familiar with their code, they could tell what their teammates had changed simply by reading through the code base. Software also provided this type of implicit cue automatically. P6, used the fact that his text editor highlighted lines of code that had been changed, but not yet committed to the codebase, to review what he had edited in his last programming session.

Mental representation. A large portion of context was not represented explicitly or implicitly in external artifacts, but was held in memory. This was especially the case for information regarding people, connections between aspects of context, and the history of the activity. Some goals, particularly process goals and constraints, were represented mentally such as the time constraints in the following description of what P3 was thinking upon resuming a graphic design task:

"She [the administrative assistant] sent a notice out... saying Dr. Smith is giving the next talk in this series. And I saw it and was like 'That's in a month and I need to make the poster because it takes two weeks to get the poster to the framer and they frame a poster for the speaker.'"

This description demonstrates how explicit cues, implicit cues, and memory are entangled. Whereas the details of Dr. Smith's talk were explicit in the email, the fact that the talk deadline served as a deadline for P3 to complete the seminar poster was implied. Being new to her job, the administrative assistant who had sent the announcement email did not know P3 was in charge of designing the poster, or that

the seminar even required a poster. Interpreting the poster deadline from the email thus required P3's particular knowledge. Moreover, the specific steps and estimated time needed to design and print the poster were not explicit in the email, or even closely implied by it, but were drawn from P3's memory as she considered whether she needed to respond to the email right away or could wait a few days before acting.

Strategies

Even when explicit and implicit cues are present in artifacts, finding and making sense of them can take substantial effort. We observed participants using a broad range of resumption strategies that involved key artifacts, organizing artifacts, and templates.

Review key artifacts. Participants spent time reviewing key artifacts before proceeding. This included interactions such as clicking through and briefly scanning all open windows (e.g., P2 clicked through his IDE, terminal, file browser, and web browser before editing code), reading through a single document in depth (e.g., P10 read through several sections of her paper before editing one of them), or simply re-reading the last few lines of a document before proceeding (e.g., P4 briefly read an incomplete paragraph before continuing to edit it). This method seemed to be used most often to search for implicit context about goals, plans, actions, and artifacts. However, sometimes participants encountered explicit context while reading through key artifacts, such as when P6 came across a teammate's inline comment about a bug or when P10 found a comment about needing to clarify her description of a particular topic in her qualifying paper.

Edit key artifacts. Participants also reconstructed context by making small edits to key artifacts. For example, P4 and P9 both started writing sessions by adding previously read papers to their bibliographies. Making these small edits reminded them of what artifacts they could reference while writing and extended the time they had to search memory for relevant context before resuming the main writing task. Programmers similarly made small edits to CSS files before making larger edits to Javascript code.

Review organizing artifacts. In some cases, participants reviewed organizing artifacts such as todo lists and paper notes while resuming their work. This strategy seemed to work best for deciding which goal to pursue next, or for reviewing a list of artifacts or information relevant to the task at hand. As opposed to finding implicit cues while reviewing key artifacts, participants most often found explicit context while reviewing organizing artifacts. P9 reviewed a bulleted summary of news articles she had compiled earlier when deciding what to add to her thesis chapter. P1 was able to see not only unaccomplished goals as she reviewed her team's todo list in Figure 3, but also the history of goals she and her teammates had completed.

Edit organizing artifacts. Participants also began many resumption episodes by editing organizing artifacts (e.g., adding items to their todo lists or writing notes to themselves). These edits helped externalize mental context so it could be manipulated, evaluated, or offloaded from internal memory. For example, P6 was able to look at a list of programming goals he had just created and compare their priority when deciding

which one to pursue next. Moreover putting a goal down on a todo list seemed to help some participants to then focus on the task at hand.

Organizing artifacts often included information about multiple activities. P4 started writing todo items for one project and then began to write items for another project before returning to start working on the first project. Editing organizing artifacts also let context be more easily shared with collaborators.

Review a template. In a number of cases, participants opened a template to structure their work. For example, when crafting the seminar poster, P3 pulled up a poster she had created for a previous seminar. Not only did this save time that would have been devoted to creating and styling redundant elements, visually scanning the template poster provided a way to implicitly determine what goals she needed to achieve (e.g., update the title, modify the color scheme, and so on). P10 searched for prior meeting notes to use as a template when drafting a handout for an upcoming meeting. And all of the programmers we observed looked for example code to serve as templates for specific functions they needed to write, holding implicit cues of goals, plans, or actions that could be used to achieve a particular goal.

4.3. Discussion

Activities structure human actions in a cohesive manner. While activities can be planned, aimed at goals, and composed of actions enacted with artifacts, our observations reiterate the recurring finding in HCI that even highly directed activities such as programming, writing, and graphic design are not simple linear processes unfolding over time. Rather, they involve the consideration and selection of possible actions in response to dynamic material and social situations. In short, activities are situated actions (Suchman, 1986).

Reconstruction as Situated Action

The process of reconstructing mental context for these activities is also composed of situated actions. As opposed to past characterizations of reconstructing context as a process of restoring suspended goals, plans, actions, and the artifacts needed to enact them, our observations of everyday knowledge work demonstrate how the process involves reconstruction and consideration of a host of other information about the activity's context. This includes information about collaborators, the history of the activity, and connections between these and goals, plans, actions, artifacts, and their attributes.

This process of reconstructing a network of contextual information appears inefficient when viewed from the classic information processing perspective in which the historical model is a suspended computer routine that only needs to be resumed from its location in a call-stack hierarchy. In this view, context reinstatement is akin to finding a vector of where one was and where they were headed. However, such an approach breaks down when the activity is not highly procedural or well defined, as is the case with most

knowledge work. Knowing one's last action does not necessarily reveal what needs to be done next. Moreover, knowledge workers frequently need to reprioritize tasks, reset goals, and generate new plans when resuming activities. Our participants had to place the activities they were resuming in a temporal perspective, specifically figuring out past goals, plans, actions, and artifacts, how they had evolved, and anticipating changes in the future. This process of dynamically deciding what needs to be done next requires a broader understanding of the activity.

The environment plays an important role in this cognitive reconstruction. While explicit reminders and rote memory play a part, our observations highlight the prevalence of using implicit cues pulled from the environment to reconstruct context. These cues are ambiguous and commonly a natural byproduct of prior activity. Moreover, they often require additional understanding of the structure and history of the activity to interpret. Implicit cues are useful for a number of reasons. Workers do not need to intentionally create them, as they do with handwritten notes and other explicit cues. This frees them from the metacognitive task of organizing and explaining ill-formed work as they go. Implicit cues are also typically found in key artifacts, the very artifacts needed to enact an activity. This reduces the chances of creating and then forgetting an organizing artifact, such as a paper outline, or failing to recognize what activity the cue refers to, like a cryptic note saying "check the accounts".

The social world also plays a vital role in reinstating context. Even activities that are traditionally thought of as solitary, such as writing a thesis, involve a significant amount of collaboration with others. The collaborative nature of pursuing activities and reconstructing their context has at least three impacts. First, mental context involves information about collaborators, their skills, knowledge, and intentions. Second, collaborators serve as a source of mental context, able to both provide information and help interpret it. Third, collaborators act on the shared material environment, changing both the state of the activity and the shared artifacts. This prevents workers from being able to assume the activity remains as they left it, and also impacts their ability to find and interpret cues in the material world.

While psychological research has begun to more explicitly account for the role of the environment in cuing mental context (Trafton, Altmann & Brock, 2005) and the process of reconstructing rather than immediately recalling mental context (Salvucci, 2010), it still often focuses on recalling a single goal or simple problem state from a clear environmental cue, such as a cursor location. However, everyday activities are often collaborative situated actions that involve richer mental context whose reconstruction involves subtle environmental cues and interactions with collaborators.

Reconstruction as Sensemaking

Possessing a network of information describing an activity is not sufficient context for resuming that activity. The information must also be interpreted and understood. Within HCI, this process of sensemaking has been described as collecting, organizing, and creating representations of information to understand and pursue problems (Russell, Stefik, Pirolli, & Card, 1993). In organizational studies, Weick (1995) has described

sensemaking as a process of making order when current situations seem unordered or different than expected. Across domains, sensemaking research has consistently found that storytelling, or constructing narratives, enables understanding. Orr (1986, 1996) highlighted how photocopier technicians used stories to diagnose machine problems and share knowledge with co-workers. In another seminal study of information needs, Gorman (1995) found that physicians used stories to communicate information and ask questions about patients.

We similarly find that stories enable people to organize and make sense of their activities and are vital to the process of reconstructing mental context. Just as narratives provide a structure for remembering and communicating information to collaborators, they can also provide a structure for communicating context from the past to the present for an individual. These stories do not seem fully formed at the start of resumption, but unfolded as participants first read an explicit note, noticed an implicit cue, or recalled a bit of context from memory and then slowly built up a surrounding network of mental context. We refer to this process as *situated sensemaking*.

Dependencies on Cognitive and Collaborative Structure

Writing, web development, and graphic design vary in numerous ways that are not captured by the data we collected for this study. However, we did note ways in which the cognitive and collaborative structure of these different activities influenced how people went about reconstructing mental context. Below, we focus on four differences in the activities: the execution-evaluation loop, relationships between artifacts, the explicitness of goals, and the use of artifacts in collaboration.

Tightness of execution-evaluation loop. One significant difference between the activities we observed from graphic design, to writing, and web development is the tightness of the execution-evaluation loop. In graphic design, actions are often taken directly on graphic elements and their effects can be evaluated immediately. In writing, evaluating a block of written text often involves scanning that text and other related portions of the document. In web development, when participants wanted to see the effects of their edits, they first had to execute the code and then sift through a network of files to determine how the change they made interacted with other portions of the codebase to produce the observed effect. As interactions got less direct, participants had to navigate between more artifacts to assess the current state of the activity and spent more time reading and editing them to recall relationships between artifacts. As a result of a looser execution-evaluation loop, we observed programmers and writers spending more time reading and making small edits to key artifacts before resuming their work than did graphic designers.

Type and visibility of relationships between parts of artifacts. Another significant difference between activities was the way in which parts of each artifact relate to one another. In graphic design, many of the relationships are visual or spatial. Participants could quickly scan the area surrounding an element they wanted to edit to see related elements or what impact their edit would have on the rest of the document. In writing, relationships between parts of a paper can be nonlinear and are often at an abstract level

rather than being directly available visually or spatially. Thus assessing the state of a document or the relationship between components often takes a significant amount of reading rather than a quick glance. In web development, the relationship between functions and blocks of code can be even more complex. Functions may call several other functions in the same file or across files. For our participants, as relationships between parts of the key artifacts grew less visio-spatial and more distributed or conceptual, they also grew more opaque. Consequently, participants doing writing or web development spent more time reading or editing key artifacts to recall relationships between artifacts. While web developers had numerous tools at their disposal for doing so (e.g., print statements, web inspectors, IDEs), writers resorted to reading their papers in detail or creating "reverse" outlines to remind themselves of the structure.

Explicitness of goals. Web development, writing, and graphic design also differ in the explicitness of their goals and measures of success. In graphic design and writing, many of the goals are implied or hard to evaluate (e.g., lucid writing, credible sources, cohesive color scheme). In programming, more goals are explicit and easily tested (e.g., create a form for registering a new customer). Goals that are explicit are easier to externalize with organizing artifacts such as todo lists. As a result, we observed web developers making more use of todo lists than writers or graphic designers, who more often recalled goals by reviewing or editing their key artifacts. Whereas all the web developers we observed created todo lists for their individual projects, none of the graphic designers we observed did. P4, a writer, wrote large goals (e.g., "revise chapter 4") on a todo list and was only able to articulate smaller sub-goals after she had been writing for a while.

It is also worth noting that while organizing artifacts were often used to store large goals, even the web developers kept smaller goals either in their minds or more flexible artifacts such as a handwritten scratch pad. As P2 mentioned while looking at his teams' shared development plan, "I have my own development plan that is not online... The development plan that I use for my own stuff is super detailed... it's something that changes a lot whereas I don't want to go and change stuff a lot [on the shared plan]."

Use of artifacts in collaboration. A final difference in the activities we observed was the use of artifacts in collaboration. With graphic design and writing, collaborators often worked on different artifacts or very distinct portions of the project. P7, a graphic designer, had a collaborator compile and send her a list of talk descriptions in a Google Doc, which she transcribed into a separate InDesign document to create a conference program. P3, another graphic designer, received photos and text from collaborators while making a seminar poster. Not only did these graphic designers work on different files than their collaborators, they also had a well defined division of labor, being in charge of layout but not content. As a result, graphic designers often needed to gather artifacts received from collaborators when resuming their work but, once these were in place, could work independently and focus on editing their key artifacts.

In web design however, collaborators worked on overlapping goals or artifacts more frequently and had to do more to check the state of the activity before they resumed work. They would check shared Github repositories to see if their teammates had

changed files since they last touched the project, review code to look for differences by comparing what they read with what they remembered the code looking like, and reference shared todo lists to look for explicit statements of whether their teammates had achieved certain goals.

5. IMPLICATIONS FOR DESIGN AND RESEARCH

Managing fragmented work is cognitively demanding, particularly because the mental context needed to resume suspended activities fades from memory over time. The above studies demonstrate that computational systems have potential to help users reconstruct the mental context of suspended activities by curating lightweight visual histories. Some computational systems already present subtle visual cues about their history of use. For example, P6 used the highlighting in his IDE to determine what lines of code he had edited in his last working session. However, more can be done to track and present visual histories, both within and across applications. Our work has several implications for the design of such systems and points to several avenues for ongoing research.

5.1. Implications for Design

Prior research has demonstrated the effectiveness of images and video in helping people remember suspended activities. More targeted studies have found that programming aids that show recent code edits rather than lists of recently edited files are more effective in helping people identify prior activities. Together these studies demonstrate that concrete, visual representations of prior work are more effective than abstract representations in cueing memory. However, there has been little guidance for designing visual histories for activities beyond programming, or activities that span multiple applications. Here we distill general design principles from our studies.

Show graphic representations of activity, not just abstractions. Participants in our second observational study made significant use of implicit cues in their working artifacts to reconstruct mental context, particularly for relationships between artifacts, relationships between goals, plans and actions, and recalling what action they intended to take next. Resumption aids should not only focus on creating abstract summaries but focus on making these implicit cues more available by showing images of prior work, leveraging people's remarkable visual memory to recall related context.

Highlight connections. The connections between artifacts, goals, plans, and actions were valuable aspects of mental context for participants to recall. Resumption aids should seek to make these connections easier to reconstruct, particularly by showing which artifacts or portions of artifacts were interacted with in succession. For example, cycling between a Stack Overflow page and a particular method of code reveals a relationship between the two artifacts as well as the plan articulated on the Stack Overflow page and the actions taken to write that block of code. Integrating with established planning tools such as todo-lists, calendars or tasks managers could help reconstruction of connections.

Highlight what has changed. Resumption aids should highlight what has changed about artifacts, goals, plans, and actions. While there are numerous tools for programmers

to review changes to their code, there are fewer tools for graphic designers and writers to review changes to their work. Highlighting changes can be especially useful for collaborative activities when workers need to understand how the activity has evolved while they were away.

Surface implicit goals and plans. Writing, graphic design, and web development all have numerous implicit goals and plans that guide the activity. These goals may be stable (e.g., make the website look professional) or change significantly over time (e.g., such as when a plan to use Google Drive as a database is abandoned). Resumption aids should seek to make these implicit goals and plans more visible or easily deduced from implicit cues. For example, showing where a number of actions were performed and then later undone reveals a plan that was attempted and then discarded.

Use full screen thumbnails, for now. Without a better predictor of users' attention than the mouse (e.g., gaze; or perhaps recent mouse activity such as clicking on an object), full-screen thumbnails can be smaller than cropped thumbnails and evoke as detailed memories. Our results also suggest that thumbnails still work for recognizing cross-application activities, but need to be larger (320px high for 80% recognition) than those used for recognizing websites (208px high according to Kaasten, Greenberg, & Edwards, 2002). This result is encouraging since a website thumbnail is a summary of one window, but a desktop thumbnail contains more content (overlapping windows, toolbars, etc.). This size increase may also be an artifact of computer monitors increasing in pixel density over the last decade. For example, the 17" screen used in the Kaasten study was 864 pixels tall whereas our participant's average screen height was over 1200 pixels. Correcting for screen height, Kaasten et al.'s website thumbnails had to be 24% of the original screen height before participants recognized the website 80% of the time whereas our desktop thumbnails had to be 26% of the original screen height before participants similarly recognized their activity.

Videos are less effective than expected. We found that videos, at least 5x time-lapses, cue less accurate memories than screenshots and found no evidence that they produce memories that are any more detailed. Showing activity over time may be better accomplished with other methods such as small-multiples, thumbnails that allow scrubbing, or tapestries.

No need for familiarity. We did not find evidence that random experience sampling and reviewing thumbnails improved memory in our tasks. This is in line with Sellen et al.'s (2007) finding that images taken passively by a wearable camera were as good of cues for remembering past events as actively taken images. This does not diminish the value of bookmarking past events to create landmarks for future reviewing, but one should not expect these events to be more memorable than others randomly selected.

5.2. Implications for Research

Prior research consistently demonstrated that people recall more of their past activities when viewing images or videos of those activities than when relying on textual descriptions or unaided memory. The two studies above build on this result by providing

evidence that people can also reconstruct rich mental context when reviewing images of past work, even when those images are relatively small. They also extend the notion of mental context beyond suspended goals, plans, and actions, to include information about collaborators, and the activity's history over time. These studies reveal that restoring mental context is a process of *situated sensemaking* rather than simply retrieving a suspended goal. However, there are many alternatives to the simple desktop thumbnail visual histories we investigated and room for more research on tailoring visual histories to the unique nature of mental context required for various activities. We encourage further research in the following areas.

Activity segmentation. In our first study we randomly sampled images of participants' computer-mediated activity. A comprehensive visual history, however, will want to parsimoniously summarize longer periods of activity. One approach would be to detect specific types of activities as well as their boundaries and identify a small number of representative visual cues for each activity. There are several potential ways to do this motivated by past research. One would be to build on prior video abstraction literature (Truong & Venkatesh, 2007), though there are challenges applying techniques that rely on video of real-world environments to video of desktop activity. A second approach is to detect activity boundaries using computer events, as Iqbal and Bailey did to identify breaks between document editing, image manipulation, and programming activity (2007). A third approach would be to augment desktop-event models of breakpoints with data from office and wearable sensors, as Fogarty et al. did in their study of interruptability (2005).

Cue selection. Another avenue for further research is in selecting visual cues to represent each activity or sub-activity. In our first study we tested thumbnails that were naïvely cropped around the user's cursor, assuming the cursor to be a proxy for their focus. This approach did not work very well and thumbnails cropped in this way had to be larger than miniaturized full screen thumbnails before participants could recognize their activity. Since sometimes people move the cursor out of the way, it would be interesting to use the history of movement as a factor to possible identify an area of the screen involved in recent activity. Future research could also explore eye tracking, saliency detection, or decomposition of the screen components (Dixon & Fogarty, 2010) to automatically identify regions of interest.

Domain specific studies of mental context. Programming specific studies has led to the development of useful tools, such as Mylyn (Kersten, 2006) for helping maintain mental context. Future studies could focus on domain-specific tools for other activities to aid restoring of mental context. For example, as compared with tools to aid programming there is a lack of computational support for the cognitive challenges of writing and rewriting.

Designing effective visual representations of activities. Visualizing past activities in an efficient manner to reconstruct mental states is particularly challenging. Even with visual cues selected to represent significant moments, there are a number of ways to arrange and make these cues interactive that could aid review and activity resumption. For example, would showing a series of small multiples be more effective than viewing a

time-lapse of images, or perhaps would a network graphs better reveal relationships between thumbnails (Waldner, Bruckner & Viola, 2014)? We have outlined some directions in the implications for design. However, more research is needed to clarify how visual histories can cue the complex and fragile network of thoughts that evolve throughout users' activities.

Preserving privacy. Creating visual histories involves recording information about user's personal and potentially confidential activities. Users may have also have unique concerns when recording work activity as opposed to personal activity. Tracking across applications may also be felt as quite invasive, even when logs are only stored on the device and kept private. We have conducted preliminary investigations following a value-sensitive design framework to better understand knowledge workers' privacy concerns regarding recording and sharing work-related computer mediated activity. These have revealed that workers consider their productivity and reputation when tracking and sharing their work-related activities. However, further research is needed to identify the multiple aspects of users privacy concerns and methods for addressing them, especially for collaborative activities where there may be substantial benefit to sharing histories.

6. CONCLUSION

Resuming interrupted activities is cognitively challenging, requiring the reconstruction of prior mental states. Surprisingly few systems have been explicitly designed to help users reinstate this mental context, though many have been developed to help users manage the timing of interruptions and to restore their digital workspaces after an interruption. To date, studies of restoring context have primarily focused on programming rather than other knowledge work activities. As a result resumption aids have been designed primarily to benefit programmers and their specific tasks.

Through two studies we demonstrate that computers might augment users ability to restore past mental context by curating lightweight visual histories of their activity. Moreover we examine how the information, artifacts, and strategies knowledge workers currently use to restore mental context vary across activity. Reconstructing mental context can be characterized as a process of *situated sensemaking* enacted by storytelling and use of explicit and implicit cues. Users construct complex networks of contextual information, elaborating on emerging cues and using memory and inference to craft stories to explain why the cues exist as they do as well as their implications. We believe that computers can help users tell these stories by curating visual histories of their work that prompt storytelling.

Some computational systems already present visual cues about history of use but more can be done to track and visualize past activities, both within and especially across applications. Helping users make sense of past activities is an essential step in supporting the reconstruction of past mental contexts and ultimately resuming suspended activities. Understanding what to visualize has implications for the nature of data that needs to be recorded, transitioning from storing only the state of artifacts to additionally capturing the history of their evolution and providing it in ways that can engage visual and episodic memory. With information technology now embedded in virtually every device and

system, understanding and managing interruptions, their impacts, and designing systems to aid with resuming suspended activities are crucial challenges for the future.

Figure 1. Resuming a complex activity often requires reconstructing mental context, which can be challenging.
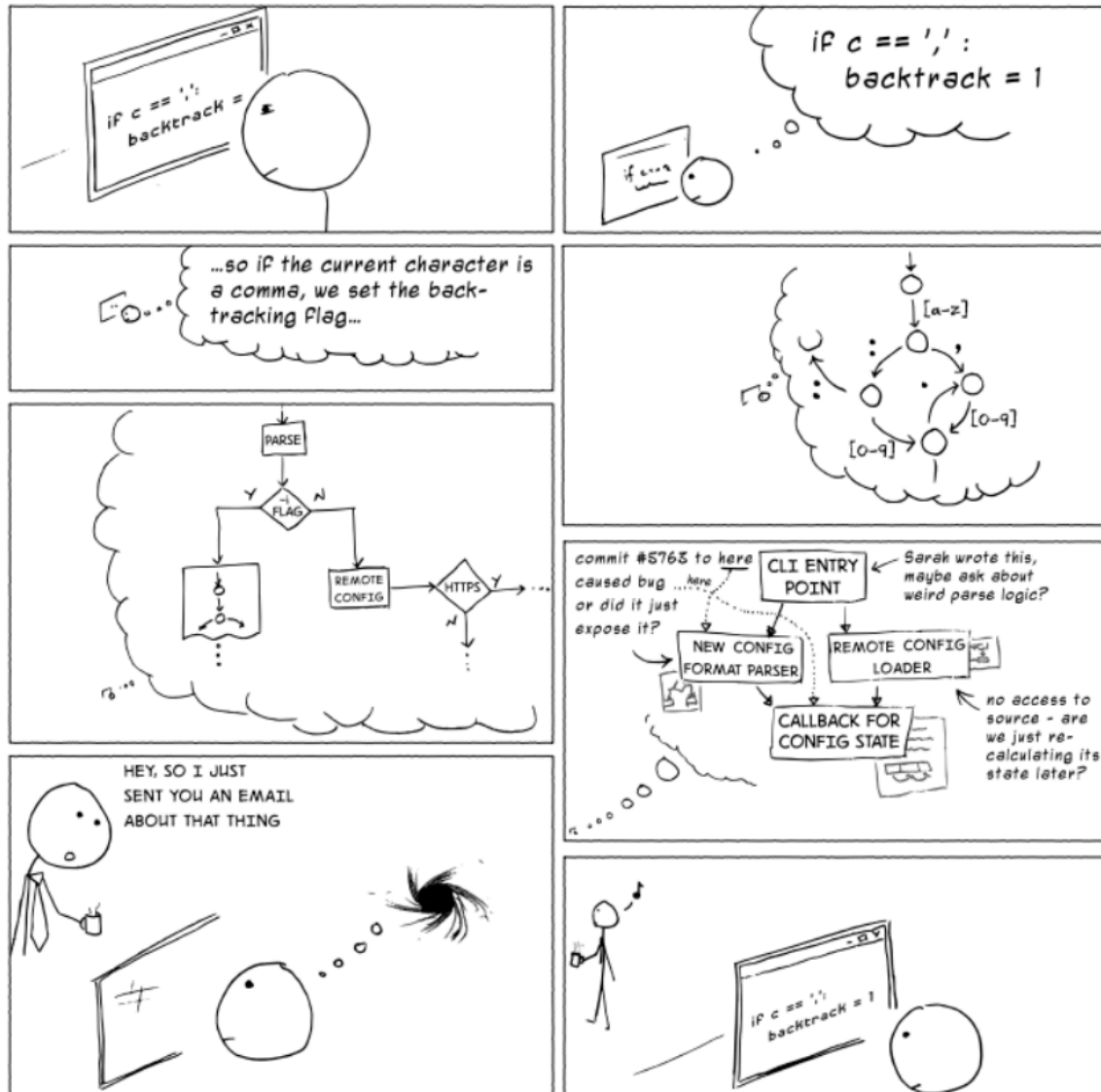
Figure 2. Cropped thumbnails (B) can be the same size as full-screen thumbnails (A) and show less of the screen in greater detail.
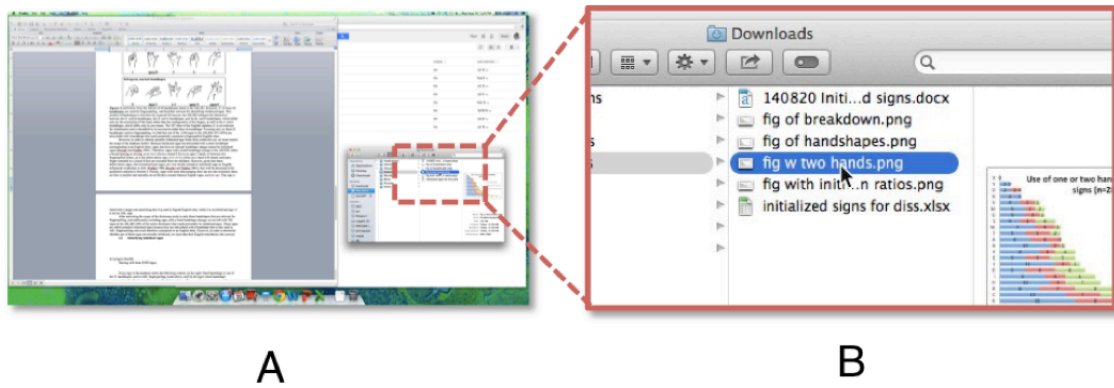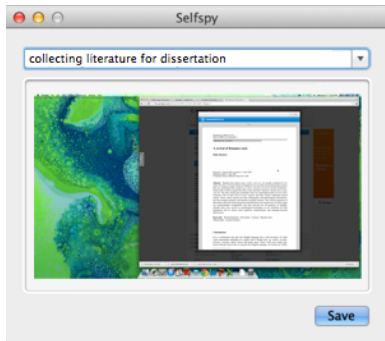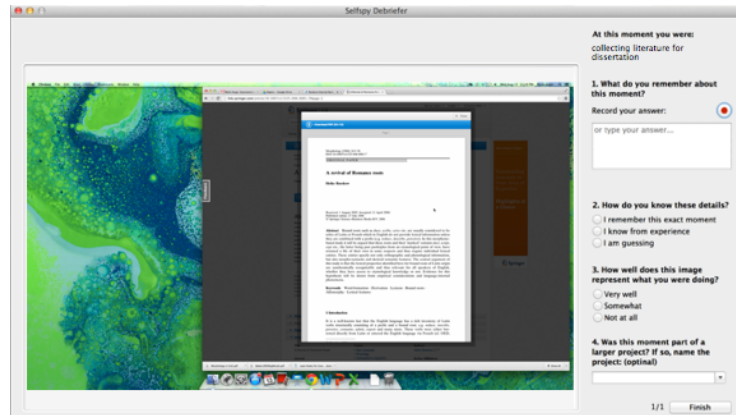

A

B

Figure 3: Participants' activities were randomly sampled with an experience-sampling window (A) and mental context collected during end of day debriefs (B).

A

B

Figure 4. Participants described what, why, who, and when they were doing an activity in a large proportion of responses.
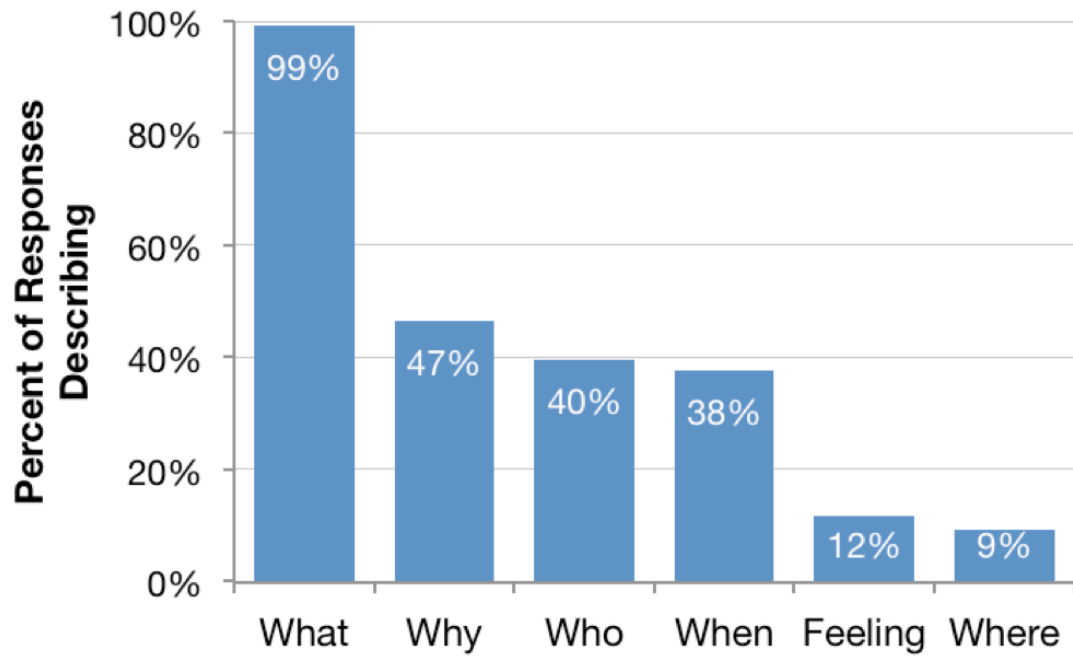
Figure 5: Example of cropped thumbnail image that one participant used to recognize a prior activity and reconstruct mental context.
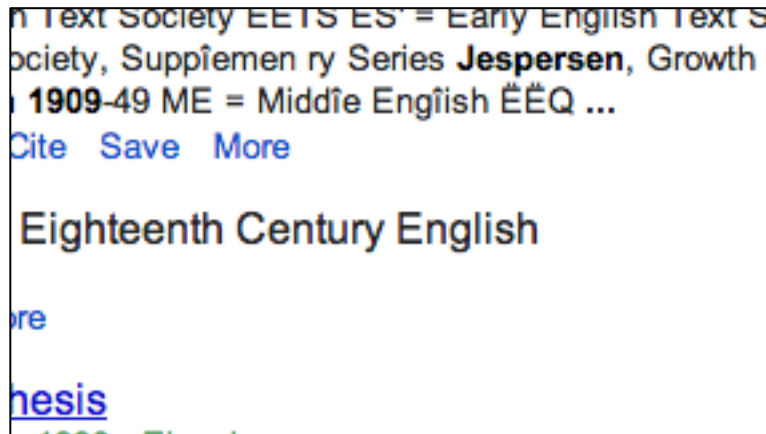
Figure 6. Participants used a ChronoViz (Fouse, Weibel, Hutchins, Hollan, 2011) visualization showing their screen recording, general level of computer activity (green line), and textual notes (red bars) to find resumption episodes. ChronoViz is an application that allows multiple streams of time-coded data to be displayed simultaneously.
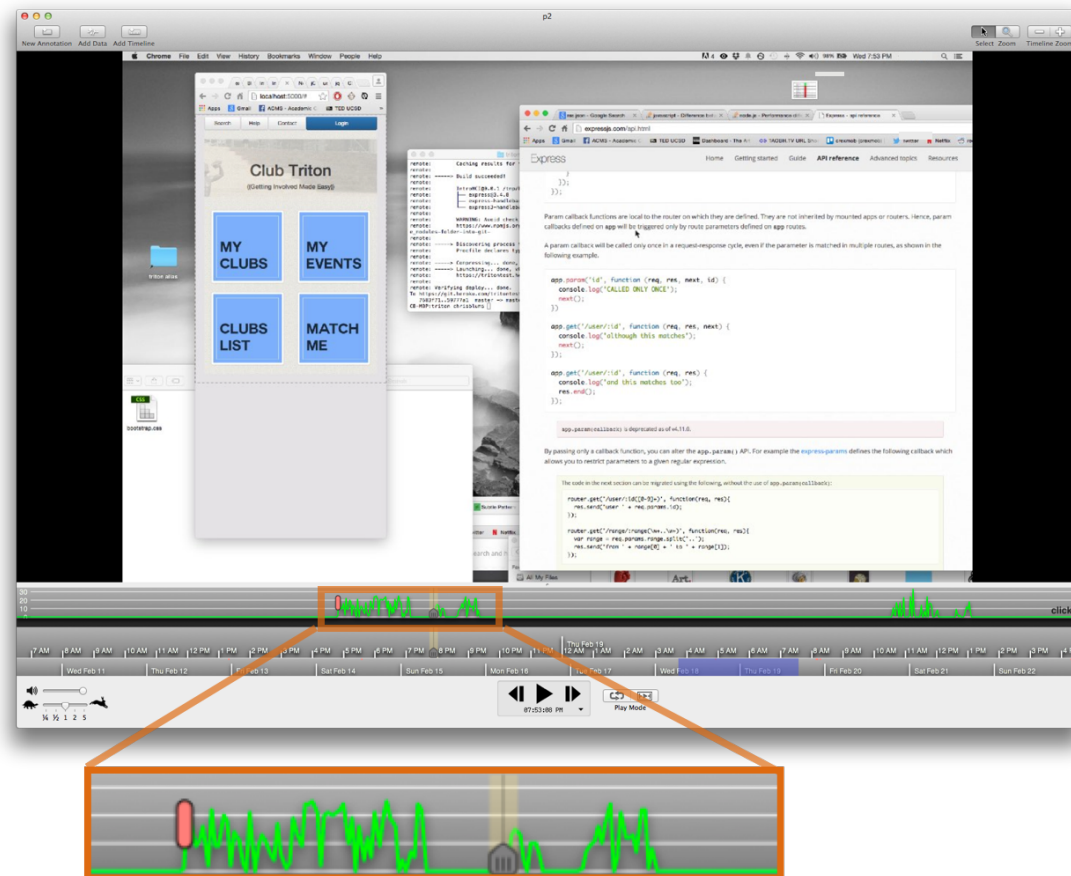
Figure 7. Explicit context was often stored in todo lists such as this shared list for a web development team.

Week	Task	Deadline
5	Create 'Log In' page	11:59 Feb 5
	Add 'Log In' button to all pages if user not logged in	11:59 Feb 5
	Create 'Upload Notes' page	11:59 Feb 5
6	Add Department on Course page, Course on Professor, etc.	11:59 Feb 12
	Remove 'Forward' button	11:59 Feb 12
	Add password strength security check when creating account	11:59 Feb 12
	Create 'Upload by...' selection page with one type of upload functional	11:59 Feb 12
	Create logged in user home page	11:59 Feb 12
	Move project to use handlebars and databases	11:59 Feb 12
7	Set up note database	11:59 Feb 19
	Set up log in database	11:59 Feb 19

Figure 8. P1 used two implicit cues in this document – the white space after "Task to Try" and the pink marker indicating a collaborator's cursor location – to decide what to do next.
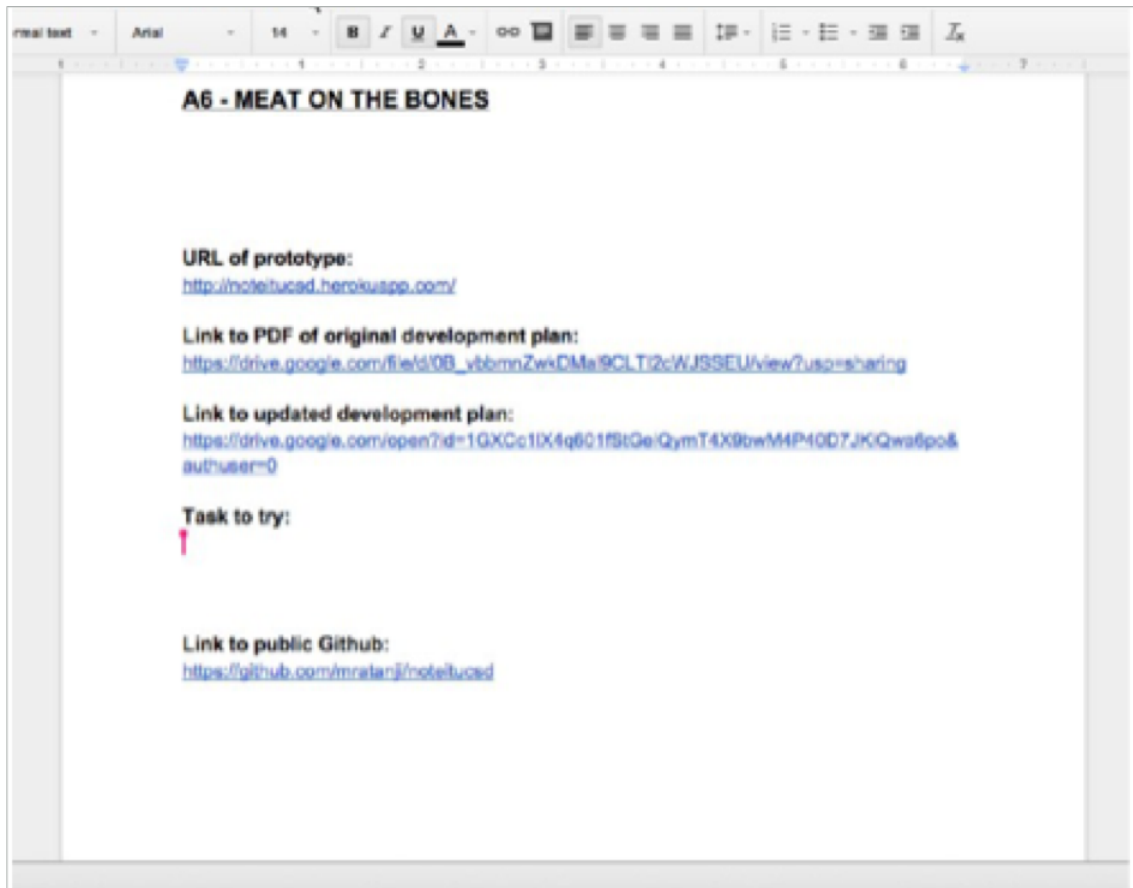
NOTES

<div align="center">**References**</div>

Adamczyk, P. D., & Bailey, B. P. (2004). If not now, when?: the effects of interruption at different moments within task execution. *Proceedings of the CHI 2004 conference on Human Factors in Computer Systems*. New York: ACM.

Altmann, E. M., & Trafton, J. G. (2002). Memory for goals: An activation-based model. *Cognitive Science*, 26(1), 39–83.

Bardram, J., Bunde-Pedersen, J., & Soegaard, M. (2006). Support for activity-based computing in a personal computing operating system. *Proceedings of the CHI 2006 conference on Human Factors in Computer Systems*. New York: ACM.

Bernstein, M. S., Shrager, J., & Winograd, T. (2008). Taskposé: exploring fluid boundaries in an associative window visualization. *Proceedings of the UIST 2008 Symposium on User Interface Software and Technology*. New York: ACM.

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences*, 105(38), 14325–14329.

Brandt, J., Weiss, N., & Klemmer, S. R. (2007). txt 4 l8r: lowering the burden for diary studies under mobile conditions. *Proceedings of the CHI 2007 conference on Human Factors in Computer Systems*. New York: ACM.

Cangiano, G. R., & Hollan, J. D. (2009). Capturing and restoring the context of everyday work: A case study at a law office. *Proceedings of the 2009 international conference on Human Centered Design* (pp. 945-954). Berlin: Springer.

Czerwinski, M., & Horvitz, E. (2002). An investigation of memory for daily computing events. In Faulkner, X., Finlay, J., & Detienne, F. (Eds.), *People and Computers XVI-Memorable Yet Invisible* (pp. 229–245). London: Springer.

Czerwinski, M., Horvitz, E., & Wilhite, S. (2004). A diary study of task switching and interruptions. *Proceedings of CHI 2004 conference on Human Factors in Computer Systems*. New York: ACM.

Dabbish, L., Mark, G., & González, V. M. (2011). Why do I keep interrupting myself?: environment, habit and self-interruption. *Proceedings of the CHI 2011 conference on Human Factors in Computer Systems*. New York: ACM.

Dixon, M., & Fogarty, J. (2010). Prefab: implementing advanced behaviors using pixel-based reverse engineering of interface structure. *Proceedings of the CHI 2010 Conference on Human Factors in Computing Systems*. New York: ACM.

Dragunov, A. N., Dietterich, T. G., Johnsrude, K., McLaughlin, M., Li, L., & Herlocker,

J. L. (2005). TaskTracer: a desktop environment to support multi-tasking knowledge workers. *Proceedings of the IUI 2005 international conference on Intelligent User Interfaces*. New York: ACM.

Dumais, S., Cutrell, E., Cadiz, J. J., Jancke, G., Sarin, R., & Robbins, D. C. (2003). Stuff I've seen: a system for personal information retrieval and re-use. *Proceedings of the 2003 SIGIR conference on Research and Development in Information Retrieval.* New York: ACM.

Ellis, J., & Kvavilashvili, L. (2000). Prospective memory in 2000: Past, present, and future directions. *Applied Cognitive Psychology*, 14(7), S1–S9.

Fogarty, J., Hudson, S. E., & Lai, J. (2004). Examining the robustness of sensor-based statistical models of human interruptibility. *Proceedings of the CHI 2004 conference on Human Factors in Computer Systems*. New York: ACM.

Foroughi, C. K., Werner, N. E., Nelson, E. T., & Boehm-Davis, D. A. (2014). Do interruptions affect quality of work? *Human Factors*, 56(7), 1262–1271.

Fouse, A., Weibel, N., Hutchins, E., & Hollan, J. D. (2011). ChronoViz: a system for supporting navigation of time-coded data. *Proceedings of the CHI 2011 conference on Human Factors in Computer Systems*. New York: ACM.

Gillie, T., & Broadbent, D. (1989). What makes interruptions disruptive? A study of length, similarity, and complexity. *Psychological Research*, 50(4), 243–250.

González, V. M., & Mark, G. (2004). Constant, constant, multi-tasking craziness: managing multiple working spheres. *Proceedings of the CHI 2004 conference on Human Factors in Computer Systems*. New York: ACM.

Gorman, P. N. (1995). Information needs of physicians. *Journal of the American Society for Information Science*, 46(10), 729-736.

Grabler, F., Agrawala, M., Li, W., Dontcheva, M., & Igarashi, T. (2009). Generating photo manipulation tutorials by demonstration. *ACM Transactions on Graphics*, 28(3), 66.

Grossman, T., Matejka, J., & Fitzmaurice, G. (2010). Chronicle: capture, exploration, and playback of document workflow histories. *Proceedings of the UIST 2010 Symposium on User Interface Software and Technology*. New York: ACM.

Guo, P. J., & Seltzer, M. (2012). BURRITO: Wrapping Your Lab Notebook in Computational Infrastructure. *Proceedings of the USENIX 2012 Workshop on the Theory and Practice of Provenance*.

Hailpern, J., Jitkoff, N., Warr, A., Karahalios, K., Sesek, R., & Shkrob, N. (2011). YouPivot: improving recall with contextual search. *Proceedings of the CHI 2011 conference on Human Factors in Computer Systems*. New York: ACM.

Heer, J., Mackinlay, J., Stolte, C., & Agrawala, M. (2008). Graphical histories for visualization: Supporting analysis, communication, and evaluation. *IEEE Transactions on Visualization and Computer Graphics*, 14(6), 1189–1196.

Henderson Jr, D. A., & Card, S. (1986). Rooms: the use of multiple virtual workspaces to reduce space contention in a window-based graphical user interface. *ACM Transactions on Graphics*, 5(3), 211–243.

Hess, S. M., & Detweiler, M. C. (1994). Training to reduce the disruptive effects of interruptions. *Proceedings of the HFES 1994 Human Factors and Ergonomics Society Annual Meeting*. Sage.

Hincapié-Ramos, J. D., Voida, S., & Mark, G. (2011). A design space analysis of availability-sharing systems. *Proceedings of the UIST 2011 Symposium on User Interface Software and Technology*. New York: ACM.

Ho, J., & Intille, S. S. (2005). Using context-aware computing to reduce the perceived burden of interruptions from mobile devices. *Proceedings of the CHI 2005 conference on Human Factors in Computer Systems*. New York: ACM.

Hodges, S., Williams, L., Berry, E., Izadi, S., Srinivasan, J., Butler, A., Smyth, G., Kapur, N., & Wood, K. (2006). SenseCam: A retrospective memory aid. *Proceedings of the UbiComp 2006 International Conference on Ubiquitous Computing*. Springer.

Horvitz, E., Koch, P., & Apacible, J. (2004). BusyBody: creating and fielding personalized models of the cost of interruption. *Proceedings of the CSCW 2004 Conference on Computer Supported Cooperative Work*. New York: ACM.

Houben, S., Bardram, J. E., Vermeulen, J., Luyten, K., & Coninx, K. (2013). Activity-centric support for ad hoc knowledge work: a case study of co-activity manager. *Proceedings of the CHI 2013 conference on Human Factors in Computer Systems*. New York: ACM.

Hupbach, A., Hardt, O., Gomez, R., & Nadel, L. (2008). The dynamics of memory: Context-dependent updating. *Learning & Memory*, 15(8), 574–579.

Iqbal, S. T., & Bailey, B. P. (2006). Leveraging characteristics of task structure to predict the cost of interruption. *Proceedings of the CHI 2006 conference on Human Factors in Computer Systems*. New York: ACM.

Iqbal, S. T., & Bailey, B. P. (2007). Understanding and developing models for detecting and differentiating breakpoints during interactive tasks. *Proceedings of the CHI 2007 conference on Human Factors in Computer Systems*. New York: ACM.

Iqbal, S. T., & Horvitz, E. (2007). Disruption and recovery of computing tasks: field study, analysis, and directions. *Proceedings of the CHI 2007 conference on Human Factors in Computer Systems*. New York: ACM.

Jo, J., Kim, B., & Seo, J. (2015). EyeBookmark: Assisting Recovery from Interruption during Reading *Proceedings of the CHI 2015 conference on Human Factors in Computer Systems*. New York: ACM.

Jeuris, S., Houben, S., and Bardram, J. (2014). Laevo: a temporal desktop interface for integrated knowledge work. *Proceedings of the UIST 2014 Symposium on User Interface Software and Technology*. New York: ACM.

Kaasten, S., Greenberg, S., & Edwards, C. (2001). *How people recognize previously seen Web pages from titles, URLs. and Thumbnails*. Tech Report 2001-692-15, Department of Computer Science, University of Calgary, Alberta, Canada.

Karger, D. R., Bakshi, K., Huynh, D., Quan, D., & Sinha, V. (2005). Haystack: A customizable general-purpose information management tool for end users of semistructured data. *Proceedings of the CIDR 2005 conference on Innovative Data Systems Research*.

Kaptelinin, V. (2003). UMEA: translating interaction histories into project contexts. *Proceedings of the CHI 2003 conference on Human Factors in Computer Systems*. New York: ACM.

Kersten, M., Elves, R., & Murphy, G. C. (2006). WYSIWYN: Using Task Focus to Ease Collaboration. *Supporting the Social Side of Large Scale Software Development*, 19.

Lamming, M. G., & Newman, W. M. (1992). Activity-based information retrieval technology in support of personal memory. In *IFIP Congress (3)* (Vol. 14, pp. 68–81). Citeseer.

Lee, M. L., & Dey, A. K. (2008). Lifelogging memory appliance for people with episodic memory impairment. *Proceedings of the UBICOMP 2008 conference on Ubiquitous Computing*. NewYork: ACM.

Lottridge, D. M., Rosakranse, C., Oh, C. S., Westwood, S. J., Baldoni, K. A., Mann, A. S., & Nass, C. I. (2015). The Effects of Chronic Multitasking on Analytical Writing. *Proceedings of the CHI 2015 conference on Human Factors in Computer Systems*. New York: ACM.

Mark, G., Gonzalez, V. M., & Harris, J. (2005). No task left behind?: examining the nature of fragmented work. *Proceedings of the CHI 2005 conference on Human Factors in Computer Systems*. New York: ACM.

Mark, G., Gudith, D., & Klocke, U. (2008). The cost of interrupted work: more speed and stress. *Proceedings of the CHI 2008 conference on Human Factors in Computer Systems*. New York: ACM.

Mark, G., Iqbal, S., Czerwinski, M., & Johns, P. (2015). Focused, Aroused, but so Distractible: Temporal Perspectives on Multitasking and Communications.

Proceedings of the CSCW 2015 Conference on Computer Supported Cooperative Work. New York: ACM.

Mark, G., Voida, S., & Cardello, A. (2012). A pace not dictated by electrons: an empirical study of work without email. *Proceedings of the CHI 2005 conference on Human Factors in Computer Systems*. New York: ACM.

Mark, G., Wang, Y., & Niiya, M. (2014). Stress and multitasking in everyday college life: an empirical study of online activity. *Proceedings of the CHI 2014 Conference on Human Factors in Computing Systems*. New York: ACM.

McFarlane, D. C., & Latorella, K. A. (2002). The scope and importance of human interruption in human-computer interaction design. *Human-Computer Interaction*, 17(1), 1–61.

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

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

Monk, C. A., Trafton, J. G., & Boehm-Davis, D. A. (2008). The effect of interruption duration and demand on resuming suspended goals. *Journal of Experimental Psychology: Applied*, 14(4), 299–313.

Newport, C. (2016). *Deep work: Rules for focused success in a distracted world*. New York: Hachette.

O'Conaill, B., & Frohlich, D. (1995). Timespace in the workplace: Dealing with interruptions. *Proceedings of the CHI 1995 conference on Human Factors in Computer Systems*. New York: ACM.

Orr, J. E. (1986). Narratives at work: Story telling as cooperative diagnostic activity. *Proceedings of the CSCW 1986 Conference on Computer Supported Cooperative Work*. New York: ACM.

Orr, J. E. (1996). *Talking about machines: an ethnography of a modern job*. Ithaca, N.Y: ILR Press.

Parnin, C., & DeLine, R. (2010). Evaluating cues for resuming interrupted programming tasks. *Proceedings of the CHI 2010 conference on Human Factors in Computer Systems*. New York: ACM.

Parnin, C., & Rugaber, S. (2011). Resumption strategies for interrupted programming tasks. *Software Quality Journal*, 19(1), 5–34.

Parnin, C., & Rugaber, S. (2012). Programmer information needs after memory failure. *Proceedings of ICPC 2012 international conference on Program Comprehension*. IEEE.

Pollan, M. (1997). *A place of my own*. London: Bloomsbury.

Rattenbury, T., & Canny, J. (2007). CAAD: an automatic task support system. *Proceedings of the CHI 2007 conference on Human Factors in Computer Systems*. New York: ACM.

Robertson, G., Horvitz, E., Czerwinski, M., Baudisch, P., Hutchings, D. R., Meyers, B., Robbins, D., & Smith, G. (2004). Scalable Fabric: flexible task management. *Proceedings of the AVI 2004 working conference on Advanced Visual Interfaces*. New York: ACM.

Russell, D. M., Stefik, M. J., Pirolli, P., & Card, S. K. (1993). The cost structure of sensemaking. *Proceedings of the CHI 1993 conference on Human Factors in Computer Systems*. New York: ACM.

Safer, I., & Murphy, G. C. (2007). Comparing episodic and semantic interfaces for task boundary identification. *Proceedings CASCON 2007 conference of the Center for Advanced Studies on Collaborative Research*. IBM Corp.

Salvucci, D. D. (2010). On reconstruction of task context after interruption. In *Proceedings of the CHI 2010 conference on Human Factors in Computer Systems*. New York: ACM.

Sellen, A. J., Fogg, A., Aitken, M., Hodges, S., Rother, C., & Wood, K. (2007). Do life-logging technologies support memory for the past?: an experimental study using sensecam. *Proceedings of the CHI 2007 conference on Human Factors in Computer Systems*. New York: ACM.

Seltman, H. J. (2012). Experimental design and analysis. Online at: http://www.stat.cmu.edu/hseltman/309/Book/Book.

Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychonomic Science*, 19(2), 73-74.

Suchman, L. (1986). *Plans and situated actions*. New York: Cambridge University.

Tabard, A., Mackay, W. E., & Eastmond, E. (2008). From individual to collaborative: the evolution of prism, a hybrid laboratory notebook. *Proceedings of the CSCW 2008 Conference on Computer Supported Cooperative Work*. New York: ACM

Trafton, J. G., Altmann, E. M., & Brock, D. P. (2005). Huh, what was I doing? How people use environmental cues after an interruption. *Proceedings of the 2005 Human Factors and Ergonomics Society Annual Meeting*. Sage.

Trafton, J. G., Altmann, E. M., Brock, D. P., & Mintz, F. E. (2003). Preparing to resume an interrupted task: effects of prospective goal encoding and retrospective rehearsal. *International Journal of Human-Computer Studies*, 58(5), 583–603.

Truong, B. T., & Venkatesh, S. (2007). Video abstraction: A systematic review and classification. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 3(1), 3–37.

Van Solingen, R., Berghout, E., & van Latum, F. (1998). Interrupts: just a minute never is. *IEEE Software*, 15(5), 97–103.

Voida, S., Mynatt, E. D., & Edwards, W. K. (2008). Re-framing the desktop interface around the activities of knowledge work. *Proceedings of the UIST 2008 Symposium on User Interface Software and Technology*. New York: ACM.

Waldner, M., Bruckner, S., & Viola, I. (2014). Graphical histories of information foraging. In Proceedings of the NordiCHI 2014 Nordic Conference on Human-Computer Interaction. New York: ACM.

Weick, Karl E. (1995) *Sensemaking in organizations*. (Vol. 3). Thousand Oaks, CA: Sage.