Task, Interrupted: Understanding the Effect of Time Costs on Information Interruptions During Data Entry

Author:
Judith BORGHOUTS

Supervisors:
Dr. Duncan BRUMBY
Prof. Anna COX

March 24, 2018

UCL Interaction Centre,
Department of Psychology,
University College London

Abstract

Data entry work often involves retrieving data from one or more sources in the environment and entering it into a computer system. Though this type of self-interruption is required to complete the task, switching between looking up and entering the required data can be time-consuming and disruptive, and it can be difficult to remain focused on the task. Furthermore, the cost to access different information sources may vary, which can further influence the disruptiveness of these 'information interruptions'. Though the phenomenon of work fragmentation and interruptions is well-researched, it is unclear when, how often, and how long people self-interrupt themselves for a data entry task, and whether the type of, and access to, information sources involved influences their strategies.

This thesis investigates how information access costs (IAC) affect the number, duration, and timing of information interruptions for a data entry task. Seven studies are reported across three chapters to understand the impact of IAC in the context of entering expenses in a finance office setting.

The first part of the thesis describes two qualitative studies looking at the context in which office workers in finance offices perform data entry tasks. Interview findings from Study 1 revealed that many data entry tasks have to be scheduled over time, and a critical component of data entry work is not just entering the data, but also retrieving data from multiple sources distributed in the environment. Participants explained that they batched similar tasks to efficiently complete their work, and held items in memory whilst switching between sources. Observations in Study 2 revealed that people adopt different strategies when organising information from physical or digital sources. Physical sources take time to access and participants therefore prepare it beforehand, or postpone retrieving it until a more convenient moment in the task. Digital sources are retrieved using the same device as that where the data entry occurs, and participants often interrupt their main entry task to switch between different windows and look up this information as soon as they need it. These switches can often take longer than intended, and participants were observed being logged out of the entry system, resuming the wrong data entry task, and reported it took time to resume their work after these longer switches.

The second part of the thesis reports three lab experiments that further test the influence of information access costs on people's information retrieval strategies. These studies show that, in a controlled setting where participants can learn the time costs involved in accessing information, they first switch to information sources that are fast to access, and switch more frequently to these sources. On the other hand, people either prepare or postpone looking up information which takes time. [Study 3 showed that if people retrieve all data from the same source, they will reduce switches between entering and looking up data if the access costs to this source increases. As it took more time to access, offloading behaviour was observed as well, and several participants placed items they were going to need nearby, but did not use them yet]. Study 4 further demonstrates that when people have to manage multiple sources, they collect and group items that are quick to access first, and leave items that take longer to access until the end. Study 5 shows that this effect also applies in a multi-task setup: when dealing with two data entry tasks, people will interleave between tasks in order to enter items with a low IAC first. As a result, participants made more omission errors and submitted tasks before they had

completed entering all the items. xx

The final part of the thesis reports two studies that evaluate the effectiveness of a design intervention which aims to make information access cost more salient, and gives users explicit feedback on time spent to access information. Study 6 found that using an experimental data entry task, people who were shown how long they were away for made shorter switches, were faster to complete the task and made fewer data entry errors. Study 7 evaluated the intervention with finance workers processing expenses. Quantitative data replicated the findings from study 6 in-the-wild - participants with the intervention made shorter interruptions during the period that interruptions were logged. Data from post-study interviews indicated that time feedback made participants more aware of their switches, and they tried to remain focused on looking up information and return to the data entry task on time. They postponed interruptions until a more convenient moment in the data entry task, rather than switching often and addressing an information need as it emerged.

This thesis demonstrates how looking up information influences people's data entry work, by testing the effect of information access costs on people's switching strategies between looking up and entering data, and evaluating how making information access cost more salient can influence their behaviour. It makes a theoretical contribution by showing how people adapt to small changes in information access costs not only by changing the number and duration of switches, but also the scheduling of these switches during the main task. It makes a practical contribution by showing how making information access costs more salient influences people's switching behaviour, and can help people make their switches shorter, and schedule them at more convenient moments during a task.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Chapter outline

In this chapter, I outline the problem I am addressing in this thesis, and the proposed contribution.

Imagine, you have come back from a business trip and have to claim back your expenses. You open the expenses system and enter the first few personal details such as your name from the top of your head without any problem. For the expenses you want to claim back, you get up to collect your receipts from your wallet and start entering the items and prices into the system. The prices are in a foreign currency, so you leave the expenses system to go to a currency converter website and convert the prices. You then need to enter your account number. You go and get your wallet with bank cards from your bag. How many times have you stopped entering to go and look up certain information? Or did you get all the required information beforehand, so you did not have to get up in-between? You may have entered the information you knew first, and left the remaining items until the end? Whatever way you chose to complete the task, it involved making decisions on how to manage subtasks of looking up the required information for the entry task.

Data entry is usually a straightforward task, but when information has to be retrieved from multiple sources, not all of which are as easy to access, this simple task can quickly become complex. Switching between entering data and looking up the required data can be disruptive, as you have to pay a time cost each time you resume your data entry task. You may have forgotten where you were, enter information in the wrong place, or misremember what you were supposed to enter and enter something incorrect. This disruption can introduce data entry errors, the outcomes of which can range from annoying to more serious consequences, such as transferring money to the wrong bank account, or transferring a wrong amount.

Because it is important data entry is done accurately, data entry research has looked at improving the input interfaces (e.g. Oladimeji, Thimbleby, & Cox, 2013; Vertanen, Memmi, Emge, Reyal, & Kristensson, 2015; Wiseman, Cox, Brumby, Gould, & Carroll, 2013), but it is not clear its design implications generalise to the type of situation as described above. In this situation, you do not enter all items from a source nearby, as in most data entry studies, but need to retrieve and enter different types of information (e.g. numbers, alphanumeric strings, words) from multiple sources (e.g. e-mails, paper receipts, different computer windows). The cost to access these resources varies, and if information becomes harder to access, people increasingly rely on what they have memorised of the information rather than recheck the resource (Gray, Sims, Fu, & Schoelles, 2006). It is however not known how people prioritise and manage their tasks when they have to retrieve the information from multiple resources, with varying access costs, or how a data entry interface should be designed to support people in these situations.

While it has been shown how changing the design of a data entry interface can affect and improve how we enter data, and that different information access costs affect how often people consult one source, it is not known how the interface can support situations when data has to be retrieved and entered from various sources with varying information access costs. In order to design interactive systems that truly support this type of data entry task, it is necessary to get a detailed understanding of how users manage this task and its subtasks, and to what extent the access to the required resources affect their strategies.

This thesis investigates how different information access costs affect how people manage subtasks of looking up information for a data entry task, with the aim to inform the design of data entry systems. In order to design systems that truly support the task they are intended for, it is necessary to get a detailed understanding of this task. For the scope of this thesis, I look specifically at how employees in financial administration offices manage looking up information from various sources for an expenses task. This task requires entering different types of information from a variety of sources, such as paper, spreadsheets, emails and databases. It is important the task is done accurately but within a reasonable amount of time as they are under time pressure to finish work on time. This data entry task is therefore considered to be an appropriate and interesting example to study further.

The expenses task serves as an example of a wider class of data entry tasks, and it can be imagined findings of this thesis can be useful and generalise to other, similar, tasks. For example, people who have to fill in their tax returns have to similarly enter a range of information into a computer system, and have to collect this information from multiple sources with varying IACs.

The research questions of this thesis are:

1. Does the IAC of information sources affect how people manage subtasks to look up information for an expenses task in a finance office setting?
   (a) What (types of) information do people need for entering expenses?
   (b) Where do people need information from?
   (c) How do they address those needs?
   (d) Does the type of information, and where they need it from, affect how people address these needs?
2. How can the existing data entry system be redesigned to support or better support people managing subtasks of looking up information for entering expenses, given variable IACs for required information sources?

The first contribution of this thesis is to map out the fragmented nature of an expenses task, of which looking up information is a substantial part of the task. It investigates how the information access cost of required information sources affects how people manage when they look up certain information. This finding has implications for the design of the current system: a second contribution is demonstrating how changing certain design features can better support people in managing these subtasks of looking up information.

1.1 Thesis outline

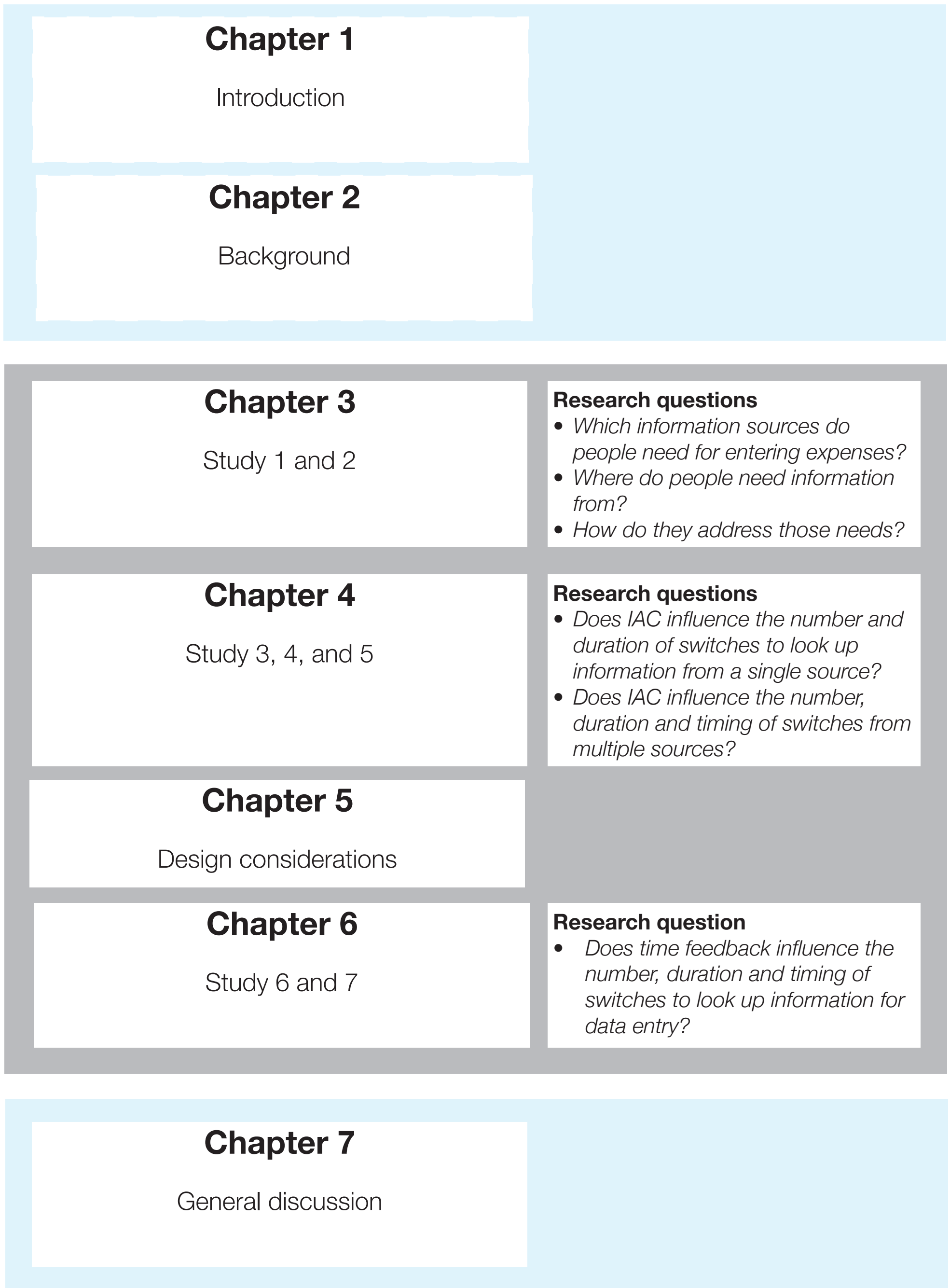


Figure 1.1: Visual overview of the thesis structure.

Chapter 2

Background

Chapter outline

In this chapter, I review previous literature on which I build my research on expenses tasks in a financial office setting. The first half of the chapter will focus on the setting, and discuss challenges of performing tasks in an office environment. The second half of the chapter will focus on the data entry task, and how certain design changes can influence people's speed and accuracy.

2.1 Disruptive environments

The previous section described the different stages of the data entry task, and how changing the data entry interface can influence people's strategies, accuracy and speed. Challenges of data entry in office settings is that data can be fragmented across documents, applications, and tasks: people deal with an increasing amount of information which they need for the task, they have to manage multiple tasks and often interrupt or get interrupted of doing their primary task. This second part of the literature review gives an overview of work that has been done to meet these challenges of working in an office environment.

2.1.1 Interruptions at the workplace

Office workers change their main work activity on average once every 11 minutes (Mark, Gonzalez, & Harris, 2005), and reportedly switch between tasks 50 times over one work week (Czerwinski, Horvitz, & Wilhite, 2004). Users can leave the primary task interface because they are distracted by a notification or because they deliberately decide to switch to a completely other task, but can also leave to look up task-relevant information. The subtask of looking up information is relevant to the primary task so may not always be labelled as an interruption from the main activity, but resuming the primary task of entering information can still be difficult (Rule & Youngstrom, 2013).

2.1.2 Reducing interruptions: time focus applications

2.1.3 Reducing information access costs: multiple and larger screens

Most computer applications are designed to support a part of a task, but do not consider that users may have to use multiple applications for the same activity (**?**). Several studies on office workers have found that information work often does not happen within one application or document, but can be highly fragmented (Czerwinski et al., 2004; Mark et al., 2005; Sellberg & Susi, 2014; **?**). People have to go in and out of several applications (Iqbal & Horvitz, 2007; **?**), switch between tasks (Czerwinski et al., 2004; Mark et al., 2005), and have to coordinate information distributed over paper files, electronic databases, e-mails, as well as people (Sellberg & Susi, 2014).

Office workers are dealing with an increasing amount of information, and may have to leave the task interface frequently to look up information. As discussed earlier, hiding information and increasing access costs may have a positive effect, as people encode the information better in memory. They do not need to look up the information as frequently and are more efficient in completing the task (e.g. Waldron, Patrick, & Duggan, 2011). Having the information well-memorised can also make people more robust against interruptions (Morgan, Patrick, Waldron, King, & Patrick, 2009) and showing data in a harder-to-read font can make people more accurate in text and number transcription tasks (**?**).

In these studies the information to hold in memory was limited and came from one source. When dealing with a lot of information from multiple sources, decisions on what has to be presented where and in what manner is still a challenge and it is often up to users to arrange the required resources for the task (Bardram, Bunde-Pedersen, & Soegaard, 2006; Grudin, 2001).

Flipping back and forth between several documents can make you lose train of thought, and when interrupted you may have forgotten the original intention of what you intended to look up (Grudin, 2001). In order to help people manage information and minimise switching between windows, both multiple and larger screens have been introduced into the workplace. While a number of studies have favoured one large screen over two smaller ones (e.g. Bi & Balakrishnan, 2009), the two solutions seem to have different benefits.

Bi & Balakrishnan (2009) conducted a study comparing people's behaviour in office work when they had to use one small screen, two small screen, and one large screen. People arranged information on their available screen space differently in each condition. For two screens, people dedicated one screen for their primary task which filled up the entire screen. They moved all information they did not need at the moment to the second screen and did not bother re-arranging the windows: they would only attend to the second screen when they needed information, and deliberately allocated the second screen for a different purpose than the first screen. For one large screen, participants first spent a certain amount of time optimising the layout of the windows by resizing and re-arranging them. They put all windows needed for the primary task in the center of the screen and placed other windows in the periphery. As participants needed information from this periphery, they dragged the window to the center of the screen rather than interacting with that particular part of the screen. On the other hand, if participants needed information from the second screen, they physically turned to the second screen and interacted with it but did not drag the information to the primary screen, unless they had to interact with it for a longer time.

For the majority of tasks, participants preferred one large screen over two screens, as it made them feel more immersed in, and surrounded by, their work. Two screens had a physical distance between them, so large documents could not easily be seen across the screens. A further drawback from having two screens was that it needed more physical space on a desk. Bi & Balakrishnan (2009) mentioned that a large screen made people better able to monitor other applications during work, such as emails. This was viewed by the authors as a benefit, but may also be more distracting. If such applications are instead placed on a second screen, people will be less disrupted by them because their visual attention is not there, but they can still easily glance at it to see if there are any new notifications (Grudin, 2001).

Despite the popularity of one large screen, Grudin (2001) states dividing screen space up in multiple monitors can sometimes be better. He argued that the main benefit of having a second screen is not so much the increase in screen space, but the partitioning of information into dedicated areas. He compares screen size with a house: sometimes it is better to have two rooms rather than one big room, as you can use the rooms for different purposes, such as one for a bedroom and one for a living room. Similarly, having multiple screens prompts people to think more about where to put which information. To support his argument, he conducted a field study looking at how office workers use multiple screens to arrange information. Participants positioned information they did not need at the moment on a second screen where they were not distracted by it, but could easily access it when needed. People preferred that information was always in the same known location and referred to the second screen whenever they needed to look information up, even when they were aware they could also access it using their primary screen as well, where the information was sometimes less time-consuming to access (Grudin, 2001).

A larger screen can better support people in multitasking, as people will not need to flick back and forth between windows as often, but can have these open and on the screen simultaneously (Czerwinski et al., 2003). However, it can also have a snowball effect: as screen size increases, so does the number of open windows, and people engage in more complex multitasking (Robertson et al., 2005). In addition, a large screen may reduce having to click and tab between windows but it can introduce another type of access cost, which Robertson et al. (2005) refer to as the *'distal information access cost problem'*: as screen size increases, it becomes harder and more time-consuming to target and select certain buttons and windows.

2.1.4 Managing information needs

Increased screen space allows people to have more windows open at once. This may reduce the need for people to open and close windows and hold information in memory when flicking between windows, but the responsibility stills lies with the user to first collect and organise all the required resources for the task (Bardram et al., 2006). Some task management applications have been built to support users to collect and group information needed for the task, but people often do not know the complete context of their activity yet and the information they will need. In addition, manually categorising and grouping task-related information can be time-consuming (**?**).

People do not always realise they need certain information until they have started a task. Leaving the primary task interface to look up this information does not have to be bad if it is useful for the current activity, but resuming the primary task after this interruption still takes time (Rule & Youngstrom, 2013). Sohn, Li, Griswold, & Hollan (2008) conducted a diary study to get an insight in people's information needs on mobile phones and how they address these needs. They did not focus on a particular task, but asked participants to keep a record of all instances where they needed information for an activity they were doing.

The information needed was retrieved from both public and personal sources on their mobile phones, such as website and e-mails, as well as physical locations. Four main factors determined whether participants looked up the information when they needed it, when they addressed it later or when they did not address it at all: urgency, importance, cost, and situational factors. The more urgent and important it was to have the information for the activity they were doing, the more likely it was they looked up the information at the moment they realised they needed it. The more time or monetary cost was associated with getting the information, the more likely they were to not address it or leave it until later. Other reasons for looking it up later or not at all, was that they were currently involved in an activity that made it difficult to address the information need at that moment, or they did not know where to get the information from.

People sometimes had to switch between several information sources and applications for one task, such as personal e-mails, search engines, and Google Maps. The more effort and time it took to go in and out of these several sources, the more likely it was that the participant gave up on looking up the information at that moment and deferred it until later. The source where people needed to get the information from, and how costly it was to access this source, also influenced whether people looked up information as they needed it or not.

The authors conclude mobile devices should take into account the activity people are doing to make sure information related to a task becomes easy to access. Though it may be difficult to predict which information will be needed, it should be at least easier for users to switch between different information sources.

The information needs in this study were primarily for personal rather than work activities, and people may be less flexible in deciding to not look up information at all when it is work-related. They may however decide to look up information later, and if it is costly to access information or depending on importance of accuracy or efficiency, decide to rely on information they have memorised rather than looking up the information of an external source.

Summary

Office work can be highly fragmented, and people have to go in and out of several documents, applications and windows for the same activity. It is often up to the user to collect information, and it should be made easier for people to collect information for a task and resume it when they leave the primary task interface to look up information. The cost to access information, the urgency and importance of an information need, and the current situation people are in influences their decision on whether they look up information as they need it or not.

Office workers deal with an increasing amount of information and it is challenging to decide how to present information needed for the task. An increased screen size can reduce certain access costs, such as mouse clicks to flick back and forth between windows, but may introduce other access costs, such as time to select the right window. Multiple and larger screens both increase screen space, but people adopt different behaviour and strategies for each set-up. Two monitors makes people dedicate a second screen for all information sources and secondary tasks, so they can focus their attention on one screen. When using one large screen, people will re-arrange windows to have more windows open at the same time and focus on the primary task but still be aware of other windows.

2.2 Data entry

A data entry task can be broken down in four stages (see Figure 2.1). An information source contains the input, i.e. the data to be entered. In the perception stage, the user perceives this data. In the encoding stage, the user encodes these in the mind. In the execution stage, the user enters them into a device which produces the output of the task, namely the data entered. Additionally, there can be a checking stage where the user checks their entered output against the original input to see if it matches.

The following four sections will describe each stage of the task in turn, and will discuss research that has been done to reduce errors at this stage. Throughout this thesis the terms data, text, and number entry are used. If data is used, this refers to alphabetic, alphanumeric, and numeric characters. If the term text is used and no particular clarification is given, this refers to alphabetic text. If numbers are discussed and no specification is given, this refers to the Arabic notation of numbers, in other words digits.

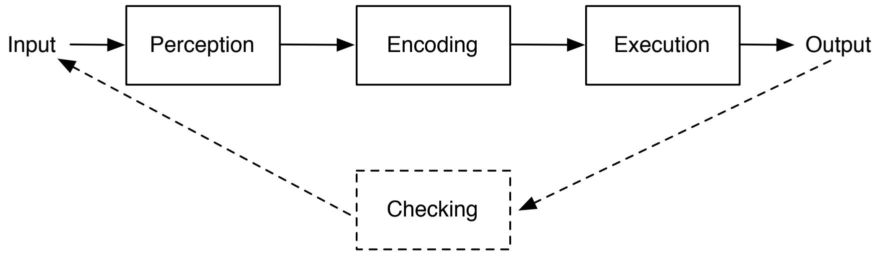


Figure 2.1: The different stages of a routine data entry task: a user perceives data input, encodes these in the mind, executes certain actions to enter data and produce the output, and can check the output against the original input.

2.2.1 The perception stage

A data entry task begins with the user looking at the data that has to be entered on a data source. Both the design of the data source as well as the data itself can ultimately have an influence on task performance. In this section, I will describe the following themes: the distribution of internal and external cognition, external representations, the memorability of data, and how presenting data in a disfluent manner can influence people's cognitive encoding of the information.

Distributed cognition

When people perform a task, they can make use of information in their mind, or retrieve information from the external environment (Norman, 1993). The distributed cognition approach has been used as a theoretical framework to explain how people make use of internal and external information to carry out work (Hollan, Hutchins, & Kirsh, 2000). In contrast with traditional cognitive science approaches that take the individual mind at the center of analysis, distributed cognition explains how the completion of a task is determined by one or more users interacting with each other, their environment, as well as external artefacts (Hutchins, 1995). The framework suggests that external representations are not merely a memory aid to off-load the limited capacity of someone's working memory, but form an integral part of cognition. In other words, it is not only the amount of external information, but also its format that can considerably affect performance (Gong & Zhang, 2009; Zhang & Wang, 2009). To be able to understand how people work, it is not enough to know how the mind processes information but it is also important to know how the information is arranged in the physical world (Hollan et al., 2000).

? argue that some proponents of distributed cognition assume people off-load as much cognition to external resources as possible. However, as will be discussed in section 2.2.2, the decisions people make on whether to offload or not is better understood as adaptive, and is affected by context as well as the design of the interface or artefacts.

External representations

Building on the work that external representations are part of cognition and that their format influence how people process and understand information, work has been carried out to explore different ways to represent information to the user.

An important factor in designing appropriate external representations is to consider how users currently use these representations. Hutchins (1995) gives a case study of cockpits as an example, in which the dial to indicate airspeed was replaced with a more precise digital display. This design did not take pilots into consideration, and only after interviews was it revealed that the new display did not match with how pilots made use of this external information. Pilots did not think of the speed as a number, but used the spatial structure of a dial to perceive their current speed and its proximity to the desired speed. By replacing the dial with digits, they lost this information.

With a better understanding of how people use representations in a certain context, these representations can also be organised in a way to make people adapt to more desirable strategies. **?** looked at programming infusion pumps in hospitals, where it is best practice to program each pump one by one, rather than interleaving in-between pumps. They found that the visual organisation of information on charts can encourage this best practice: if numeric values belonging to one pump were grouped together, participants were more likely to first finish programming one pump, before they started programming another pump.

Memorability of data

People's performance in copying data is not only influenced by how the data is organised on the source, but also by the data itself. From text entry, it is known that words are easier to transcribe than non-words as they are more meaningful and thus have a stronger representation in memory (Salthouse, 1986). Based on this, **?** investigated if the same was true for numbers. Experiments showed that familiar numbers are faster to transcribe, suggesting that these are more strongly represented in memory than random numbers as well. An applied example is PIN numbers, which is usually entered very quickly at around two seconds and at a low error rate (De Luca, Langheinrich, & Hussmann, 2010). In this case, both the digits as well as the unique motor movements made across the keypad are well encoded in memory (Mangen & Velay, 2010).

In order to make numbers in a transcription task easier to memorise and check for errors, Sandnes (2013) explored how numbers could be represented as words instead, based on the assumption that words are easier to remember and are easier to check for errors. In this paper, she presented a fixed dictionary with 100,000 numbers, each of which had a word associated with it. Instead of reading and typing the number, the user would read and type in the word that belongs to it, and a computer system would identify the corresponding number. Words would be easier to remember than a string of digits, and it would also be easier for a system to establish if an error is likely, when a word is typed that is not found in the dictionary. This novel way of presenting numbers can reduce cognitive load on the user and catch errors, but may not always be applicable depending on the domain. Numbers can be a random sequence of digits such as a phone number, but they may also have a specific meaning, such as in the financial and medical domain where a number can represent a a financial amount or medication volume. By abstracting the numbers and representing them as words that bear no relation with the meaning of the number, users do not have a sense anymore of the meaning of the numbers they are copying, which may not always be desirable.

Presenting data in a disfluent manner

The strength of information in memory can influence people's performance on copying that information, and users are both faster and more accurate in copying familiar words and numbers, rather than abstract strings of characters. If however the data to copy is abstract to the user, and cannot easily be changed or mapped onto a familiar word or other representation, there are still other ways to encourage people to encode data more deeply.

Designers of information sheets may intuitively try to present information in a way that is easy to perceive for the user, but some studies have shown that making information more difficult to perceive can have a positive effect. Diemand-Yauman, Oppenheimer, & Vaughan (2011) presented text in a grey and hard-to-read font, and found this aided people in processing and understanding the text better than when the text was presented in a normal font. The idea behind this study was that a hard-to-read font forced people to make more of an effort to read and understand the text, and as a result the text was more deeply processed and encoded in the mind. Building on this study, **?** conducted two studies where people had to transcribe text and numbers that were presented either in a black font colour or a harder-to-read grey font colour. Participants made fewer data transcription errors if data was shown in the harder-to-read font colour, both for transcribing text and numbers. There was no difference in speed between the hard-to-read and normal font colour, suggesting that the improved accuracy was not due to a speed-accuracy trade-off. This is an important finding because it shows that a deeper encoding of information may not only be beneficial when understanding text, but also when information has to be transcribed. It also shows that changing the representation can influence the encoding strategies people adopt. Further work is needed to see if this effect will remain over time, as people may either get used to the font and the effect will weaken, or people may become frustrated.

Summary

The first part of the data entry task entails the user perceiving the input. A stronger representation of data in memory can improve task performance. Data may be strongly presented in memory because the data to enter is familiar to the user, but the design of the input source can also influence people's cognitive strategies and encourage a deeper encoding. Both internal and external information affect task performance, and where and how input is presented in the external environment affects people's internal representation of it.

2.2.2 The encoding stage

After the user has perceived the input source in a data entry task, the next stage in the task sequence is the encoding stage, where data is encoded in memory. In this section, I will discuss how the cost to access information, and the effect of writing versus typing, affects how people encode information.

Information access cost

When the user perceives data, the effort involved to encode this data into memory can be influenced by the design of the data source, but also by the ease with which the source can physically be accessed in the environment. The time, physical and/or mental effort required to access information is called information access cost (IAC), and increases in the cost to access information are intrinsic to everyday tasks (Morgan et al., 2009; Waldron et al., 2011). In order to view information needed for a task, people may have to open and reopen documents, go back to a previous page, have to switch attention between their device and paper sheets, use different screens, and often have to use and retrieve information from different sources. If information is permanently available, people adopt a display-based strategy and rely on the external display as memory source. As it becomes harder to access information however, it will take too much time to retain a display-based strategy, and people will be more likely to switch to a memory-based strategy and commit more information to memory to minimise going back and forth to this external information.

This adaptive use of memory is explained by the soft constraints hypothesis, which holds that people adapt their cognitive strategies to the constraints of a task environment with the aim to optimise task completion time (Gray et al., 2006). The hypothesis states that rather than minimising cognitive resources, people try to minimise time.

To test the robustness of the hypothesis, the effect of IAC has been tested in lab experiments on several tasks such as copying tasks (e.g. Gray et al., 2006), problem solving tasks(e.g. Morgan & Patrick, 2012), and flight simulation tasks (Waldron, Patrick, Morgan, & King, 2007). A consistent finding is that people adapt their strategies to the ease with which information can be retrieved in the environment. A memory-based strategy may be faster rather than looking up the external information, but it carries the risk that the memorised items are incorrect. Gray et al. (2006) therefore recommend that the effort it costs to access information should be kept low. Similarly, **?** advise that designers of interactive devices should not rely on weak aspects of human beings, such as working memory.

However, several studies have shown that an increased IAC can have a positive effect. In problem-solving tasks, an increased IAC resulted in people taking the time to memorise task information and more planning behaviour before making any moves, which made them more efficient in completing the task (e.g. Morgan & Patrick, 2012; Morgan, Waldron, King, & Patrick, 2007).

A memory-intensive strategy can also be useful for resuming a task after an interruption. Task interruptions are known to be disruptive, because it takes time to resume the task and it can increase errors (Back, Brumby, & Cox, 2010; Brumby, Cox, Back, & Gould, 2013; Morgan et al., 2009). Morgan et al. (2009) conducted a study looking at the effect of IAC on a copying task. People had to perform the Blocks World Task (BWT), which involves copying a pattern of coloured blocks, by dragging blocks from a resource window to a target window. They manipulated the cost to access the original source which showed the pattern they had to copy. In the Low IAC condition, the pattern was permanently visible on the screen. In the Medium IAC condition, the pattern was covered by a grey mask and participants had to hover over the mask with their mouse to reveal the pattern. In the High IAC condition, there was an additional time delay before the pattern was revealed. At certain intervals, they would get interrupted and asked to do a secondary task. As IAC increased, people made fewer but longer visits to the target pattern and memorised more of the pattern. As a result, following an interruption they were faster to resume and could copy more blocks before having to revisit the target pattern.

In an experiment by Brumby et al. (2013), people had to perform a data entry task and were interrupted several times to do a secondary task. They manipulated the cost of making an error and if the cost was high, an error would cause the participant to be locked out for 10 seconds before they could resume the primary task. In this condition, people took a longer time to resume, but were less likely to make errors. A lockout made people adopt a memory-intensive strategy and take the time to remember where they were in the sequence before resuming the task.

Back, Cox, & Brumby (2012) studied the programming of two infusion pumps, and the numbers participants had to enter were situated either next to the pumps or 50 cm away. If the numbers were next to the pumps, participants saw them as individual numbers which caused them to interleave more in between the two pumps and make more errors. However, when the numbers were further away, people grouped the numbers, and first entered all numbers of one pump, and then entered the numbers of the second pump. This strategy caused them to make fewer errors.

These studies show that a memory-based strategy can be beneficial for task performance, but only if people make the effort to encode the information and therefore have the correct information in the mind. In a data entry task, it is therefore important to know what a person is reliably able to keep in short-term memory and copy correctly at once. This information can be used to design interfaces in a way that if people choose to adopt a memory-intensive strategy, they should be encouraged to check back to external information at certain times and not try to memorise too much.

The effect of writing versus typing on the level of encoding

The design of, and access to, the data source can influence how deeply people encode data in memory. In addition, the way we execute entering or transcribing the data can influence how deeply we encode it. Mangen & Velay (2010) argue there is a strong relation of cognitive processing of information and the motoric movements we have to make to input that information, and the hand movements we make to type can influence how deeply we process the information we are inputting.

Mueller & Oppenheimer (2014) looked at note-taking amongst university students, and found that students retained more information when they made notes by hand, rather than by laptop. Participants were asked to attend a talk and make notes. The talk and exams were not part of the students' curriculum and did not count towards their final grade but were set up to resemble a typical lecture and exam: the talk was followed in a lecture theatre at their university and the content was similar to what they studied. After the talk, the students then were presented with distractor tasks. Mueller & Oppenheimer (2014) conducted one study where participants were tested straight after the distractor tasks and a second study where participants were examined after one week. In both studies, participants who made notes by hand scored significantly higher on the exams. The authors support their findings with the encoding hypothesis, which holds that the processing that occurs during note-taking influences learning and retention of learning material. Participants who took notes with laptops were more inclined to take verbatim notes and made more notes, but the information was transcribed without a deeper processing of the information that was typed. Furthermore, because it took more time to write something down versus typing it, people took the time to think about what to write down and remembered this better.

While the task of taking notes during a lecture is different from data entry, some things can be taken from this study that may be applied to data entry as well. People who have to transcribe a lot of data via a conventional keyboard may switch to automatic processing in the same way as people who take notes with laptops. Though it is unfeasible to let people transcribe data by hand in most data entry situations, it does indicate how slowing people down or applying unique movements to certain characters, instead of using the same buttons that give the same haptic feedback, may promote encoding.

Summary

At the second stage of the data entry task, the user processes the input in the mind. While usually designers aim to make it easy for people and not put too much cognitive load on the user, in some cases making the user process the information more deeply and adopt a memory-intensive strategy can have a positive effect on task performance. In the case of an interruption, taking time to remember where people were in the task sequence can reduce resumption errors. Furthermore, a deep encoding of data in memory can make people more accurate in data entry.

2.2.3 The execution stage

The third stage of the data entry task is the execution stage, which is the stage where the user performs the motoric actions to enter data into a device.

? makes a distinction between two types of errors: slips and mistakes. Mistakes happen when we enter what we have in our mind accurately, but we have the wrong thing in mind. Changing the data source can influence the level of encoding users engage in to prevent these mistakes. However, even when we have the right thing in mind, we can still enter it inaccurately, which is called a slip. In order to prevent these errors from happening, work has been done on changing the entry interface. This section gives an overview of current research on both text and number entry, and how these two forms of data entry are similar and different.

Text entry

Two primary metrics in measuring people's performance on a data entry task are speed and accuracy. These metrics are commonly understood as being trade-offs: an increase in typing speed will come at the cost of errors, and a focus on entering data accurately will come at a time cost (MacKenzie & Soukoreff, 2002; Smith, Lewis, Howes, Chu, & Green, 2008). While ideally a user should perform well on both metrics, in some situations one metric can be more important than the other. In text entry, the focus has typically been on improving entry speed while remaining an acceptable accuracy.

Two popular methods in text entry to improve performance are movement minimisation and text prediction (MacKenzie & Soukoreff, 2002). With movement minimisation, the movements of the fingers to type in text are minimised. In text prediction, the system predicts the words the user intends to enter based on typed characters so far and completes the word. An alternative method is text correction, where the system waits until a user has typed a word and then tries to detect and correct errors. This relieves users from the task to check their input, and lets them concentrate on typing. Vertanen et al. (2015) created a touch screen keyboard for mobile devices called VelociTap, which used sentence-based correction instead of word-based correction. This meant people entered a full sentence before the system tried to correct input. Experiments showed that this type of text correction significantly reduced error rate while not affecting entry rate compared to a normal touch screen phone keyboard. The correction allowed users to be able to concentrate on entering text until the end of a sentence, and they were not distracted by words being predicted or corrected after each word. Furthermore, it allowed the system to take the context of words within a sentence in account and was more accurate in determining what the correct words should be.

Number entry

Text and number entry follow the same stages outlined in Figure 2.1. As with text entry, number entry involves the user looking at a number, encoding this, and entering it into a device. There are however some aspects in which number entry is different, which is why findings from text entry research cannot always be directly applied to number entry.

For instance, text prediction or correction is hard to apply to number entry because as opposed to text, there is no dictionary to determine if a number is correct or not (Wiseman et al., 2013), making it hard to predict what the user intends to type. In some settings however there are certain numbers that are inputted more often than others. Wiseman et al. (2013) looked at number entry in hospitals and found clear patterns in the numbers being entered. In a study building on this, **?** made adaptations to existing interfaces to make it easier to type in commonly used numbers in hospitals. Fewer key presses were needed to type in common numbers and as a result numbers were typed in faster, without an increase in the error rate. This shows number entry interfaces can and should be tailored to the numbers that people have to enter.

Healy, Kole, Buck-Gengle, & Bourne (2004) looked at how performance on a number entry task changes over time. They conducted an experiment in which participants were asked to enter 640 four-digit numbers, and observed both learning and fatigue-like effects. People became faster, but also increasingly less accurate in entering numbers. The authors suggest this may be because fatigued participants switch from controlled to automatic processing where they no longer deeply encode the numbers they are entering. To reduce fatigue effects, they conducted a second experiment where halfway through the experiment participants switched hands to type in the numbers. Despite this switch, the same increase in speed and errors was shown, and the

(a) A number pad. **(b)** An incremental interface.

Figure 2.2: Two different number entry interfaces tested in Oladimeji et al.'s study.

authors suggest the the major cause of fatigue in data entry was cognitive rather than motoric, as a switch in hands should be more effective to muscular fatigue than attentional fatigue. It could also be the case that people became bored and wanted to be done with the task, as there was no incentive to perform well. However, while people became faster overall in entering numbers, their initiation time increased, which means that people took a longer time at the start of each number before they started entering it. Healy et al. (2004) suggests this is another indication of cognitive fatigue. The authors recommend trainees in data entry work should be warned about losing accuracy over time, and should be instructed to respond more slowly.

The speed with which users enter data can be influenced by the design of the input method. Oladimeji, Thimbleby, & Cox (2011) compared a number keypad with an incremental interface. The two types of interfaces are shown in Figure 2.2. The number keypad is most common, and is used on calculators and phones. In this interface, each digit is assigned a button and additional buttons are usually a decimal point and a delete key to correct an error, as shown in Figure 2.2a. In an incremental interface, a number is entered by increasing or decreasing the number using up and down keys. The incremental interface used in Oladimeji et al.'s study is shown in Figure 2.2b. The double arrows increase and decrease the number by a larger amount than the single arrows.

Results of the study showed that a number keypad allowed people to enter a number more quickly than an incremental interface, but more errors were made. With the keypad, the visual attention was more on the input keys than the display. In an incremental interface, people were changing an existing value rather than entering a new value, so they had to look at the display to see how their actions changed the current value. This attention on the display may have made it more likely for them to detect errors in time. While an incremental interface may not be feasible when entering large amounts of data as it will slow users down too much, it may be preferrable over a keypad in situations where accuracy is of great importance (Thimbleby, 2011).

In order to reduce human data entry errors, attempts have also been made to eliminate the need to manually enter data altogether. This is however not always possible, and moreover can introduce different problems. Koppel, Wetterneck, Telles, & Karsh (2008) studied medication bags with a scannable bar code to enter numbers, so that the numeric values did not have to be typed by a person. They identified 31 workarounds if the system failed, for example if the bar code was missing, blurry or when the scanner did not work properly. Furthermore, even with a scanner numbers still have to be entered manually at some point, and it abstracts the numbers, which means users do not have a sense anymore of the magnitude of the numbers they are copying (Wiseman et al., 2013).

Summary

At the third stage of the data entry task, the user inputs the data into a device. Data entry has a speed-accuracy tradeoff, and being able to enter data faster often come at the cost of an increased error rate. Keyboards that slow people down may be useful in cases where accuracy is important, but may not be feasible in situations where a lot of data has to be entered. Depending on what users have to enter, shortcuts may get introduced in keyboard design so commonly used data can be entered quicker without the cost of increased errors.

2.2.4 The checking stage

Most data transcription models consider the execution stage as the final stage of a data entry task (Salthouse, 1986; **?**), but an additional stage can be a checking stage, where people review their entered output and compare it with the original target input, to see if it matches. In addition to studies trying to prevent errors before they occur, other studies have taken a different approach and looked at how a user can detect and correct errors after they have occurred.

Double data entry

A number of studies have compared two or more of the following error checking techniques in data entry: visual checking, read aloud, and double data entry (Barchard & Pace, 2011; Barchard & Verenikina, 2013; Kawado et al., 2003). In visual checking, the user visually compares the entered data with the original data. In read aloud, data from the original source is read out loud by one user and the entries are visually checked by another user. In double data entry, data is entered twice and cross-checked by the computer system.

Overall, double data entry is shown to be the most effective method in detecting errors. This method is often used in online forms when a user has to enter a password or e-mail address twice. The reasoning is that it is unlikely the same mistake is made twice. It is also not necessary for a system to know the correct input beforehand to determine if an entry error is made or not, because the two entries are cross-checked and if they do not match, an error has been made. A disadvantage of double entry is that it requires double labour and can be time-consuming. Double data entry may therefore not be feasible for each setting. For example, Kawado et al. (2003) looked at copying medical data, and in their study 104,720 data entry fields had to be entered twice by two different people. In this study, the total times for double data entry was 74.8 hours while for the read aloud, where one person said the numbers and another person entered them, it was 57.9 hours. In situations where a lot of data has to be entered, such as data entry clerk work, it may therefore not be feasible to have to enter all data twice. Moreover, it is still possible that a mistake is repeated, even when the entry is done by two different people (Nakata, 2014). In Kawado et al.'s (2003) study, 42 errors remained undetected with the double data entry method.

Visual checking

Another method to detect errors is for the users to visually check their entered data with the original data source. Visual checking is a popular method (Tu et al., 2014), but simply asking or relying on people to check is often ineffective (Barchard & Pace, 2011; Nakata, 2014; Olsen, 2008; **?**). It is particularly hard to visually check numeric data, because there is no top-down error detection: almost any order of digits is a legal number (Lin & Wu, 2014; Nakata, 2014). Users can use the context of a word to detect if it is invalid, for example when 'wrod' is typed instead of 'word'. For numbers, if 697 is typed instead of 679 this may be harder to visually spot because 697 still counts as a valid number.

Wiseman et al. (2013) looked at a number entry task, and used a checksum to detect number entry errors. This is an extra number, that is related to the other entered numbers in a way that it is sufficient to only check the checksum to determine if the input is correct, instead of checking each entry one by one. They used infusion pump parameters in their study, and users were asked to enter the medication volume to be infused and the rate of infusion. If these two numbers were entered, the computer could calculate how long the infusion would take, and this time was the third checksum number. If the checksum was generated by the computer from the two numbers entered and this checksum had to be visually checked by the user, only 36% of the errors were caught. If the third number had to be entered by the user as well, and it was checked by the computer against the other two numbers, all errors were caught but the time to complete the task was increased by 46%. These results show that visual checking was ineffective, but that asking the user to enter extra data can cause a considerable amount of time.

Olsen (2008) conducted a lab experiment in which he simulated an internet banking tool, and participants were asked to enter account numbers from a paper sheet into a computer. After participants had entered an account number, they were presented with a confirmation screen with the input, and users were asked to check their input on this screen before submitting. Participants confirmed 88 trials where they had entered an incorrect account number. In addition, in 178 trials the simulator changed people's input to another number and this incorrect number was presented on the confirmation screen. Only 5 of these 178 errors were detected and corrected. This large amount of incorrect confirmations again suggests users do not check properly, even if they are explicitly asked to do so.

Lockouts

Given the limited effectiveness of confirmation screens (Olsen, 2008; **?**), some studies have supplemented these with lockouts, where users have to wait a short period of time before they are able to confirm and submit their input.

Green et al. (2014) introduced a lockout in a patient order entry system in a hospital. Every time a physician made an order for a patient, a prompt appeared asking to double-check the identity of the patient, and the button to continue an order would be disabled for 2.5 seconds. The prompt reduced the error rate by 30%, but the total time to complete ordering tasks was lengthened by 1.5 hours on average.

Gould, Cox, Brumby, & Wiseman (2015) studied a number entry task where after each number the submit button would be disabled for a number of seconds, and a text instruction to check input appeared on the input screen. This lockout was an effective method in encouraging people to check and detect errors in a lab setting. When the study was replicated online, a short lockout made people detect errors as well but the longer the lockout duration was, the more likely people were to switch to doing other tasks, and not check anymore. This illustrates the importance of taking the task context into account, and that findings from controlled studies do not always directly translate to an applied setting. In situations where people are able to work on other tasks, a lockout has to be brief to be effective and not induce switching to other tasks (Gould et al., 2015).

Similar switching behaviour was found by Katidioti & Taatgen (2013). They conducted a lab experiment where people had to copy information and were interrupted by chat messages. Participants were free to choose when they wanted to attend to the messages. When people were locked out in the copying task and had to wait 3 seconds before they could enter the information, they often switched to the chat message, which made them forget the information to copy and slowed them down in completing the task.

Incentives

Instead of making design changes to the interface, Li, Cox, Or, & Blandford (n.d.) investigated if people can be motivated to check by incentives. They conducted a data entry experiment and manipulated the compensation people received after the experiment. In one condition, participants received extra money if they completed the task error-free. In another condition, money was deducted from their compensation if they made an error. In the control condition they received the same compensation regardless of their performance. Adding rewards and punishments significantly changed people's checking behaviour: they made more frequent and longer checks when their payment depended on their performance, and this resulted in fewer errors.

Summary

After people have entered input, they can choose to check if what they have entered is correct. A popular method is visual checking, but people are poor at doing this, even if they are instructed to do so. It is possible to have people check, but the context needs to be considered. A lockout helped people check, but when possible to switch to other tasks people used the time to spend time on other tasks.

2.3 Conclusion

Data entry follows four stages, and there are multiple strategies to complete the task, some being more accurate or efficient than others. Studies have shown that changing the design of a data entry interface can affect and improve how we enter data, and that different information access costs affect how often people visit the source from which to look up and enter data.

It is not known yet how these data entry designs can be used in an office setting, where people have to manage multiple information sources with varying information access costs. Studies on office work have shown that work can be highly fragmented, and that people may often have to go in and out of several applications to complete their task. In order to design interactive systems that truly support this type of data entry task, it is necessary to get a detailed understanding of the task in this setting, how users manage subtasks of looking up information for the data entry task, and to what extent the costs to access required resources affect their strategies.

Chapter 3

Data entry in an office setting

Chapter outline

In this chapter I describe the findings of an explorative interview study about data entry in a financial work setting. The aim of this study was to get a better understanding of the type of data entry task people at finance offices conduct, and the physical environment in which this is done. A second planned study is described, that aims to investigate the information sources finance office workers need for an expenses task, and how they currently manage subtasks of looking up information.

3.1 Study 1: Understanding data entry work in a financial office

3.1.1 Introduction

As data entry is a common task and it is important this is done both accurately and efficiently, work has been done to design and optimise data entry interfaces to support fast and accurate data entry (e.g. Oladimeji et al., 2013; Vertanen et al., 2015; Wiseman et al., 2013). However, it is not just the input method that determines efficiency and accuracy but also other aspects of the task, such as the environment within which it is conducted (Randall & Rouncefield, 2014; **?**).

Evans & Wobbrock (2012) looked if people's text entry and mouse pointing behaviour in a lab setting was comparable to how they would normally perform these inputting tasks in their everyday life. They remotely observed people's input behaviour on their personal computer, and compared this with their performance on similar tasks in a lab. Participants installed a tool on their personal computer which logged all text entry and mouse pointing behaviour they performed in one workweek. Examples of tasks that were carried out were sending personal messages to friends and browsing the web. There were no differences in uncorrected errors or text entry speed between the lab and the field, but they did find that participants corrected more errors in the lab. This study shows that people check and correct their entries more when they are in a controlled environment and are focused on the task, though the measured behaviour on people's personal computers mostly included tasks where accuracy may not have been considered important, such as sending an informal chat message to a friend.

Table 3.1: Participant information.

University	Occupation	Years of experience	Expenses	Payroll	Budgets
A	Research Services Administration	17	x		x
	Administrator	2	x		x
	Credit Controller	1	x		
	Assistant Accountant	15	x		x
	Accounts Assistant Expenses	4	x	x	
B	Payroll and Pensions Apprentice	1	x	x	
	Payroll and Pensions Assistant	10	x	x	
	Payroll Supervisor	12	x	x	
	Payroll Officer	13	x	x	

In order to support people in their data entry work, it is important to first have a better understanding of the types of data entry tasks they have to conduct, and the physical environment in which this is done. Therefore, the first study of this thesis is an explorative study. I visited and interviewed people who conduct data entry tasks as part of their daily work in the finance departments of two universities. This user group was chosen as they have a lot of data entry tasks as part of their job, and it is an area where it is important to enter data accurately, but there is also time pressure to finish work on time. Furthermore, it was an accessible user group to approach for the researcher.

Previous research has given us a good understanding of which factors may influence people's performance on a data entry task. The current study aims to study how these factors are laid out in an applied setting. Furthermore, this study gives an opportunity to see if there are additional problems that influence data entry performance, that are currently not acknowledged in existing literature.

3.1.2 Method

Participants

Nine participants (four male) took part in the study. They were employees from two public universities and their work involved receiving various requests for payment, checking the information of these requests was correct, and entering the information along with administration data into computer systems. Ages ranged from 18 to 52 (two participants wished to not disclose their age). Their level of experience differed, with some participants having just started doing this type of job and other participants working in Finance for 17 years. All but one worked full-time. Table **??** shows further demographic details of the participants. Typical tasks participants dealt with were checking and entering expense forms sent by staff and students, paying salaries and pensions, controlling research budgets, monitoring university income and expenses and entering employee information. Participants were recruited by sending invitations to opt-in mailing lists of Finance departments, and were reimbursed with a £10 Amazon voucher.

Materials

Materials that were used during the interview were a voice recorder, a paper copy of an interview script with the interview topics and guiding questions, a consent form, an information sheet for the participant and a notebook and pen to make notes. The interview script, information sheet and consent form are included in the Appendix. Each interview covered four guiding topics, which are briefly described in Table 3.2. For each topic, a number of questions were written out beforehand. These questions were used as a starting point to get the participant talking and guide the interview. Based on what the participant was saying follow-up questions were asked. The audio transcription program ExpressScribe was used to transcribe the interviews. The data analysis programs Nvivo and Atlas.ti were used to analyse the data. Nvivo was used to code the interview transcripts and notes. Atlas.ti was used to complement the analysis in Nvivo and allowed to identify relations between codes.

Topic	Description
Job description	A description of the tasks that the interviewee deals with. The purpose of this topic was to start the interview easy and give the interviewee the opportunity to explain what their job entails.
Number transcription	This includes questions on when and how people typically enter numbers for work.
Environment	This topic includes people's physical work environment, and the organisation they are a part of.
Demonstration	Interviewees were asked to give a demonstration of entering data into their system. The aim of this part of the interview is to see the type of data entry tasks people have to do, and also gives a chance to see the information sources and systems people currently use.

Table 3.2: Interview topics.

Data recording

A voice recorder was used to audio record the interviews. One participant wished to not be audio recorded and one interview could not be audio recorded due to technical issues, so for these two interviews notes were taken of the answers. For the remaining seven interviews, notes were only made of observations and not the participants' answers. Notes were made with pen and paper. Photographs were made of the work environment and screenshots of the systems that the interviewees used.

Interviewing procedure

The interviews took place at the participants' workplace. For two interviews, the interviewee's office place was not suitable for talking so the interview took place in a common room nearby, and these participants showed their workplace and completed a demonstration of entering data after the interview. Participants were welcomed and informed about the study. They received a paper information sheet with the outline of the study and contact details of the researcher to keep for future reference. They were also asked to read and sign a consent form.

The interviews were semi-structured and took between 20 and 55 minutes. Each interview was reviewed afterwards, and findings sometimes fed into new questions being included or some questions being adapted in subsequent interviews.

Pilot interview

A pilot interview was conducted with an acquaintance of the researcher who worked in Finance, to test out the set-up of the study and questions. The interview took place at the participant's home, notes were taken with pen and paper, and the interview was audio-recorded using iMovie on a Macbook Pro.

Taking notes slowed down the flow of the interview: sometimes the interviewee stopped talking to give the interviewer the opportunity to finish taking notes. Furthermore, taking notes took attention away from what the interviewee was explaining: assumptions made during the interview did not seem to be accurate in later analysis. Therefore, it was decided that note taking would be kept to a minimum. Notes would only be made of observations that could not be taken from audio recordings.

The interviewee talked elaborately about manually converting different currencies, and identified this task as the main place where errors occurred. Therefore, questions were included about if interviewees dealt with foreign currencies and converting these.

Ethical considerations

The study was undertaken with ethical approval from the UCL Research Ethics Committee [Project ID Number UCLIC/1415/001/Staff Brumby/Borghouts]. At the start of each interview, participants were first informed verbally about the study. They were then given a consent form to read and sign, and were given an information sheet to keep. This information sheet contained the study information and contact details of the researcher and the project's principal researcher, should participants have any further questions after completion of the study. They were asked permission for the interview to be audio recorded. One participant wished not to be audio recorded and notes were taken instead.

Participants were informed that the data would be used for research purposes only and stored in accordance with the Data Protection Act 1998. They were also informed that their data would be anonymised and when used in a report or academic paper, their data would not be directly identifiable. Names of participants or the universities they were working at were not included in the interview notes and transcripts.

3.1.3 Results

Data analysis

After each interview or set of interviews, a first analysis took place. The audio recording was played back, notes were typed into a digital file and reviewed and the interview was transcribed verbatim. Several non-verbal cues were included in the transcription as well, such as when the interviewee laughed or sighed, as well as descriptions of when the interviewee was demonstrating something. The advantage of doing the transcription shortly after the interview was that it was still easy to remember from listening to the audio recording what was being demonstrated. Interesting findings and initial patterns that were apparent across the data were written down.

After all interviews had been transcribed, the transcriptions and notes were printed and the data was analysed using thematic analysis (Braun & Clarke, 2006). Anything in the data that was considered to be interesting was annotated by hand and labelled with an appropriate code. On reviewing the coding, some codes were grouped together in one code, additional codes were named, and similar codes were grouped under themes. For instance, an initial code was Notifications, such as e-mail notifications. During the second coding iteration, it was identified that people always talked about notifications in the context that they were interrupted (by a notification) rather than about notifications on its own. Therefore, notifications and interruptions were grouped into one code.

The themes were then reviewed, to see if they addressed the purpose of the study. The transcripts and notes were then imported into Nvivo and coded digitally. Atlas.ti was used to complement this analysis and allowed to identify relations between codes.

Themes

In total 51 codes were derived, and these were grouped into 8 main themes, which are listed and described in Table 3.3. If codes or separate quotes did not belong in a certain theme but were still considered relevant, they were grouped in the Other category.

Theme	Description	Quotations	Participants
Task characteristics	People described things that were particular to their task, for instance how they structured their task, whether they switched tasks, and how long they took to complete tasks.	129	9
Checking	People talked about checking data input as part of their job.	103	9
System	People talked about the computer system they were using to input data.	91	9
Environment	People described their environment, for instance they talked about their physical work setting, and the work culture of their organisation.	80	9
Data	People described the data they were dealing with, for instance the type and length of data items, and from which source they copied data.	75	9
Errors	People described situations where errors were made: who made them, why were they made, what were the consequences.	75	9
Strategy	People described the strategies they used to carry out their task.	54	9
Importance of accuracy and paper trails	People talked about the sensitivity of financial data which is why not all people are authorised to approve or access financial data, and the importance of a paper trail for data entries.	35	8
Other	People talked about things that did not fit into any other category but were still considered relevant, such as issues they experienced, or queries they often received.	74	9

Table 3.3: The themes, along with a description. The column Quotations indicates how many times this theme was brought up during interviews, and the column Participants indicates how many participants talked about it.

Each theme is described separately in subsections below, and is visualised in a diagram, which shows the theme's main codes and relationships between codes, as well as quotes in dotted squares to exemplify what type of quotes were grouped under this code. The numbers in parentheses indicate the number of quotes, and the number of interviewees who mentioned it. The description of each theme is accompanied with notes and quotes taken from the transcripts to further illustrate when this theme was mentioned. These serve as examples and are not all the instances of a theme. To differentiate notes from verbatim quotes, the quotes are in italics and double quotation marks. Words put in brackets are added by the researcher to make the quote more understandable for the reader, for instance if the interviewee is talking about 'it' or 'them'. The results are ordered according to the number of quotations associated with a theme, with the theme with the most quotations listed first. The only exception is the 'Other' theme which is described last.

Task

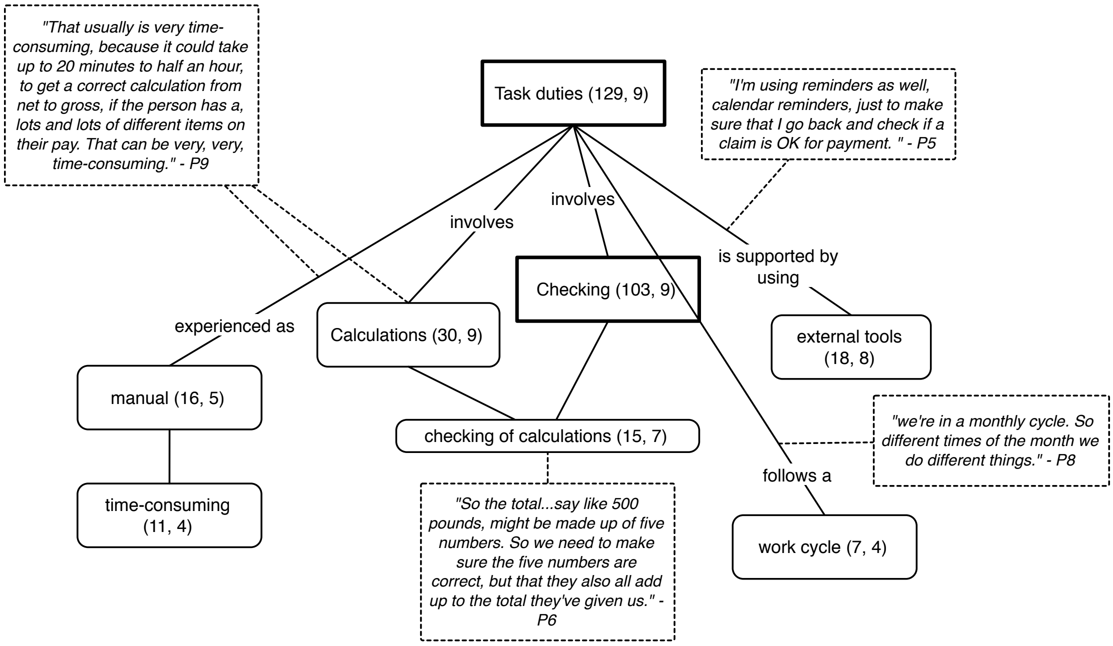


Figure 3.1: Diagram showing the theme Task. The numbers in parentheses indicate the number of quotes and the number of participants who mentioned it, respectively.

A common data entry task was entering expenses. Interviewees received expense claims from students and staff and had to check the information was correct. They then had to enter this data, along with other information such as budgetting and staff information, into a computer system. All interviewees mentioned a large part of their job was checking that data was correct. In addition to transcribing and checking individual numbers, participants also mentioned they often have to perform and check calculations. Some tasks, such as calculations, had to be done manually and this was described as time-consuming. People used several external tools in their environment to support them in their tasks. For instance, P7 and P8 had paper sheets on their desk with information they frequently had to look up, so they could easily use this to check if the input they had received was correct.

Some people worked according to a work cycle, which meant they did different things at different times of the month. One week could be reserved for checking all the data they received from another department, while another week could be spent on solely inputting data.

The time spent on each task differed: checking data on a short expenses form could be completed in two minutes, while a single calculation could take between 20 and 30 minutes. Two participants experienced dealing with a lot of data entry as tiresome.

Checking

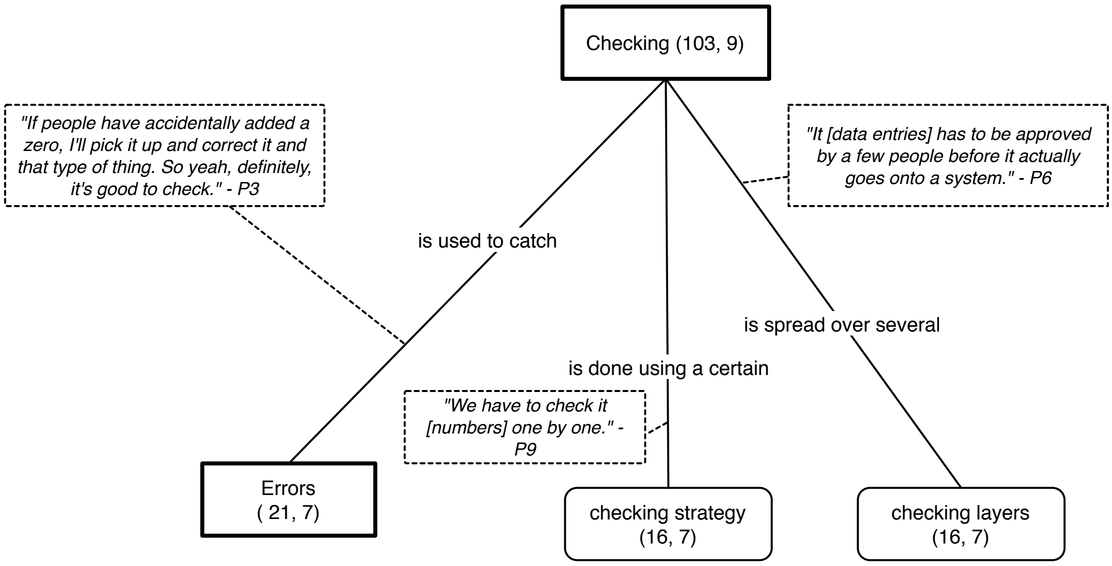


Figure 3.2: Diagram showing the theme Checking.

All participants talked about checking data input for errors as a part of their job. Data was checked by several different people and departments before it was submitted. People's experience with this checking system differed: P3 felt that an error would be caught eventually because it goes through so many different layers, whereas P9 said it made people less careful about making errors and therefore they often received erroneous data. Sending wrong input back slowed the process down, and P8, who acted as the last person in his department to check before it gets submitted, warned that even with extra human checks not all errors get caught.

People had to both check their own data input, as well as double-check other people's input. They maintained different checking strategies, ranging from checking each data item one by one to entering a group of data items first and then checking.

Participant	Quote
P3	*"we try and pick it [errors] up and then obviously there's all the different stages that pick it up as you go along."*
P9	*"the departments actually sometimes treat us as a checking system [laughs], but they shouldn't really, the schools. Because we're here just to make sure that people get paid correctly. But even though we are like a second check, we feel sometimes that we are the first checkpoint."*
P7	*"All this piece of work, when we input in the system, will be actually checked by another person... my manager will print it out, and then check... other colleagues will double-check it for you as well, the calculations."*
P8	*"one of these errors could be things that are missed during the checking."*

Table 3.4: Verbatim quotes taken from the interview transcripts that were about checking.

Participant	**Quote/note**
P1	first puts in all the details, then when done checks everything against the source.
P7	when entering numbers from paper to computer, mostly looked at paper form and the number pad; only looked at screen after finishing entering all the numbers from the form to check.
P5	*"We would go by the receipt, so we would try to make sure that the receipts are in order."*

Table 3.5: Checking own input when entering data.

Participant	**Quote**
P5	*"The numbers on the expense form will be checked individually. So the total will obviously be, say like 500 pounds, might be made up of five numbers. So we need to make sure the five numbers are correct, but that they also all add up to the total they've given us."*
P6	*"The numbers on the expense form will be checked individually."*
P9	*"We have to check it one by one."*

Table 3.6: Checking other people's input.

System

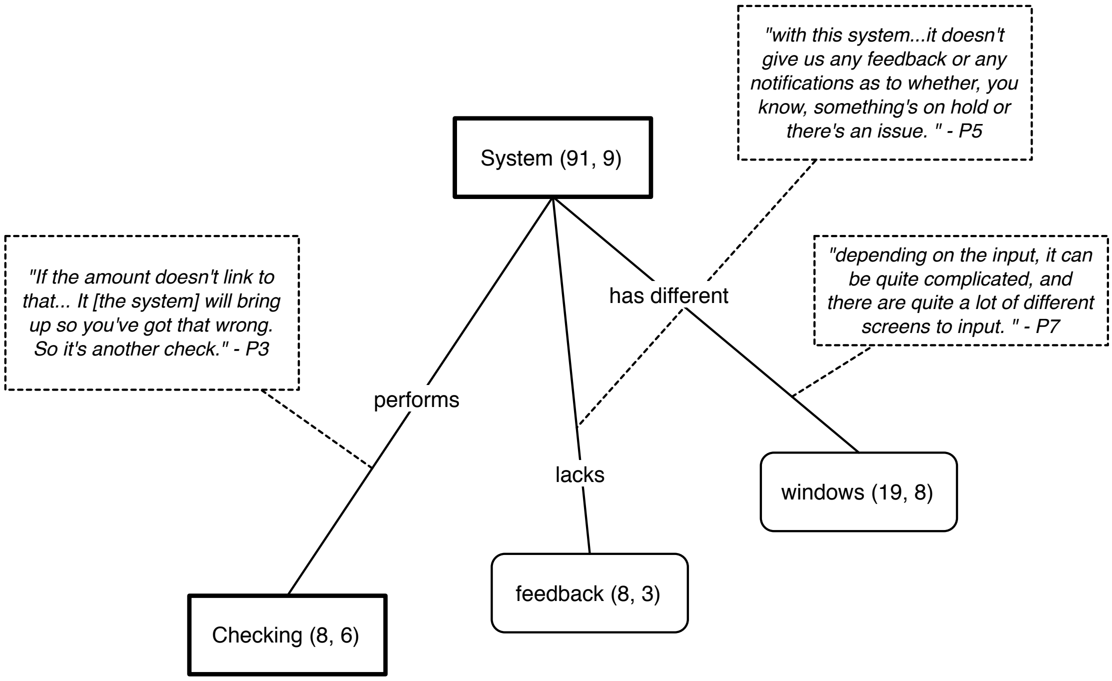


Figure 3.3: Diagram showing the theme System.

Data had to be entered into an information system on a computer. Some participants mentioned the system was another way to check for erroneous data entries: if the allowed number range of a variable was known, the system would let the user know if an out-of-range number was entered. However, it was also mentioned that one issue with the financial data entry system was that it did not give feedback if there was an issue or error.

The information needed for one task was usually spread over several windows on the computer, so sometimes people had to flick back and forth or memorise certain information from one window to use in another window.

Environment

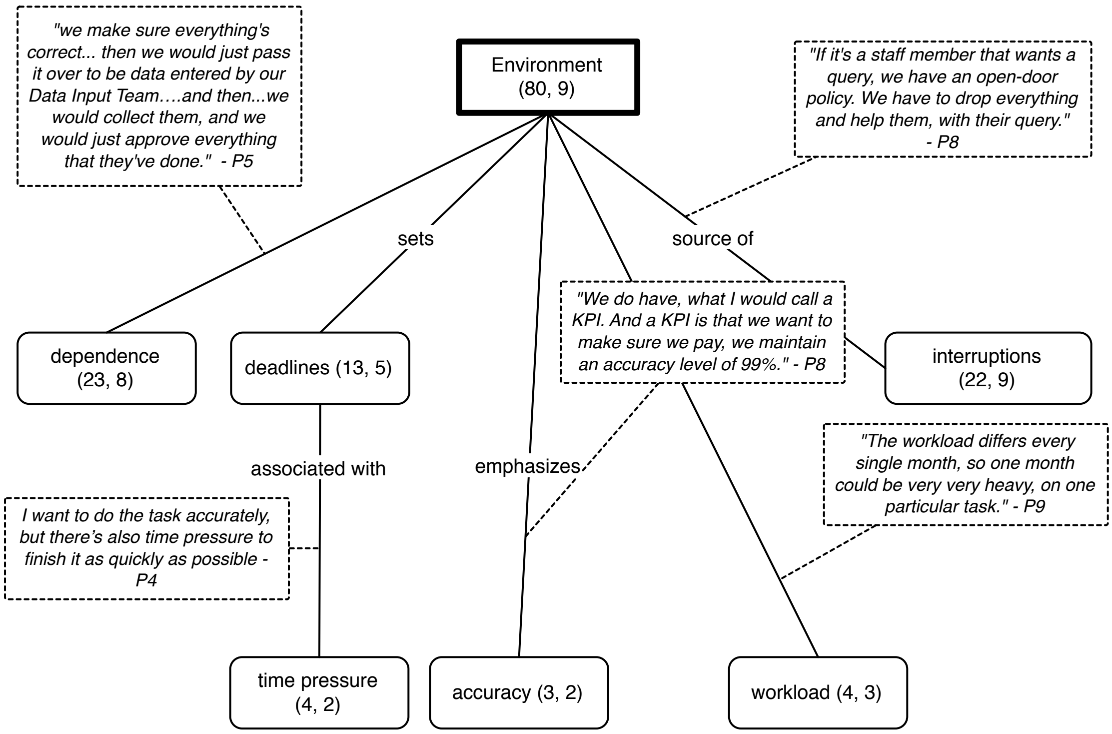


Figure 3.4: Diagram showing the theme Environment.

Every time people described the environment they were working in, this was grouped under the Environment theme. The environment could be their immediate physical setting, or their organisation in general.

All participants experienced interruptions during work. P8 mentioned he had to pause his task immediately if a staff member needed his help, but considered these interruptions part of his job. Other interviewees mentioned they tried to concentrate on the task at hand first, but did briefly attend to interruptions such as e-mail notifications, in case it was important.

People were dependent on other departments to finish a task. For instance, P5 checked paper expense forms, but did not enter these into the system herself.

Some participants felt the pressure to be both accurate in data entry, as well as finish it quickly because of deadlines. The workload varied throughout the year.

Data

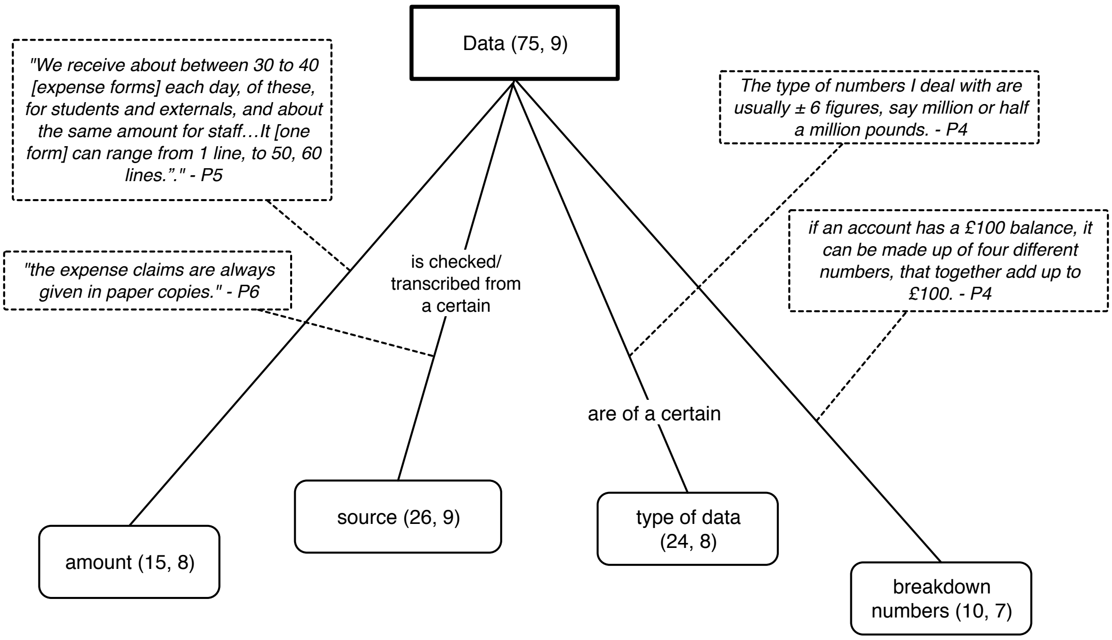


Figure 3.5: Diagram showing the theme Data.

Participants had to retrieve data from various sources. Some of the sources were electronic, such as Excel spreadsheets and Word documents. Some information had to be looked up in databases and work e-mails. Other information was received on paper sheets, and some participants had printed out information they frequently needed and had placed this on their desk. People had to transcribe numbers from paper onto a computer system for at least a part of their tasks. Some people worked with two screens. The amount of data that people dealt with differed. P5 said that for expenses alone, the amount of numbers she received each day to check and input ranges from 100 to 6000 numbers. People primarily dealt with numeric data, such as financial data and IDs. The monetary numbers they dealt with ranged from five to millions of pounds. Participants also entered and checked alphanumeric and non-numeric data such as employee names, addresses and bank account details.

Numeric data consisted of individual numbers, as well as groups of numbers that together made up a new number, such as the total amount of money spent on a project. Participants had to both check and transcribe each individual number, and check that the calculation was correct.

Errors

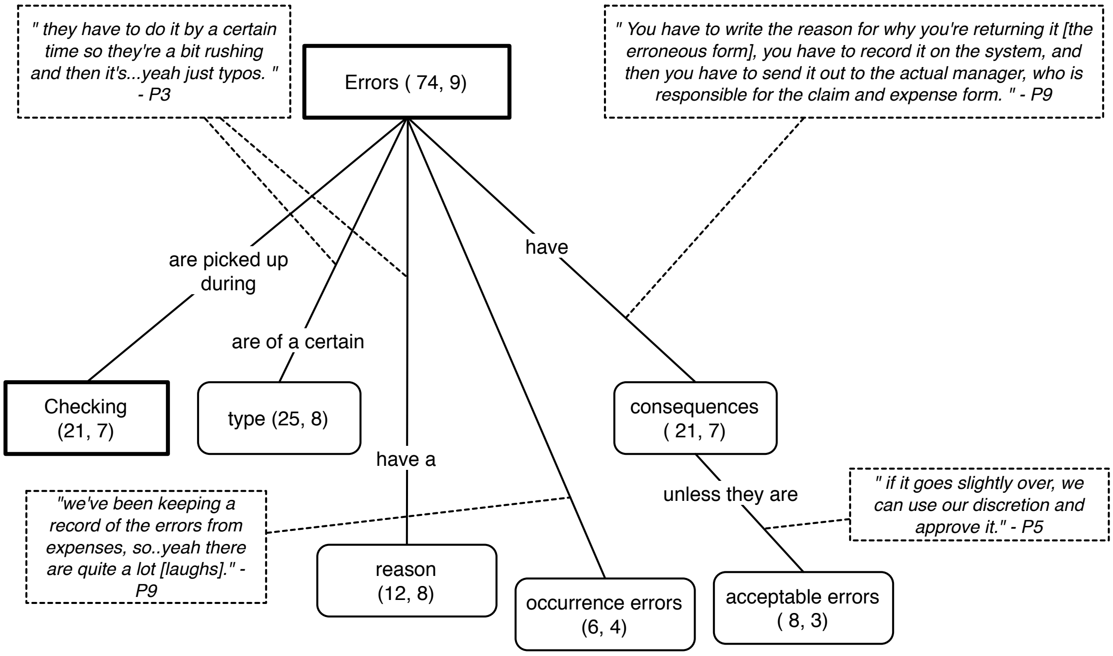


Figure 3.6: Diagram showing the theme Errors.

Participants discussed that errors happened frequently, and talked about errors they spotted from other people. Visual checking was mentioned as the main method to catch and correct errors in time, but when possible the computer system sometimes checked for erroneous data as well. Errors that were made were typos, miscalculations, or people had the wrong information to enter. For instance, P7 mentioned that when salary rates change, employees often keep entering their old salary rate on claim forms, which then has to be corrected. The main explanation people gave for errors was that it is human to make mistakes, but it was also mentioned people are under time pressure, and that people rely on the fact it will be checked by another person, which makes them less careful in entering accurate data. If an error was spotted, this had certain consequences depending on if the error was acceptable or not. If the error was sufficiently small, it could be either processed or corrected without negotiation, but if it was a large error, it had to be sent back or forwarded to a higher authority for approval.

P1 highlighted project IDs used to be letters but are now numbers, which are harder to memorise, and he felt it was easier to make a mistake. P4 worked with both paper and digital files to transcribe data from, but preferred digital files because he felt it was much easier to make an error and omit figures when transcribing from paper.

Participant	**Quote/note**
P6	*"it's quite common that we have to return an expense or payment back to someone. It happens quite often, yeah."*
P4	Yes all the time, lots of typos.

Table 3.7: People mentioned errors occur quite frequently.

Participant	Quote
P3	*"sometimes it's because people have done typos, done too many zeroes, or left out a zero."*
P5	*"the expense breakdown doesn't match what (...) whatever they put as the grand total."*

Table 3.8: The type of errors.

Participant	Quote
P9	*"Because the departments actually sometimes treat us as a checking system [laughs], but they shouldn't really."*
P7	*"Yeah, human laziness or something [laughs]."*
P8	*"sometimes, you know, through human error, you know, things don't get paid properly."*

Table 3.9: The reasons for errors.

Participant	Quote/note
P5	*"generally we tend, we try not to send claims back to departments because they might get lost in the post, and it's an inconvenience as well. So we try to... resolve it ourselves.."*
P4	We allow a certain amount of tolerance; if it turns out the thing you bought has actually decreased value and is now £40, we will allow to return £50
P7	*"we normally e-mail the budget holder to say... what you authorised is actually different. But for this kind of thing, it's only 10 pounds...we normally just process this without contacting them."*

Table 3.10: Acceptable errors.

Strategy

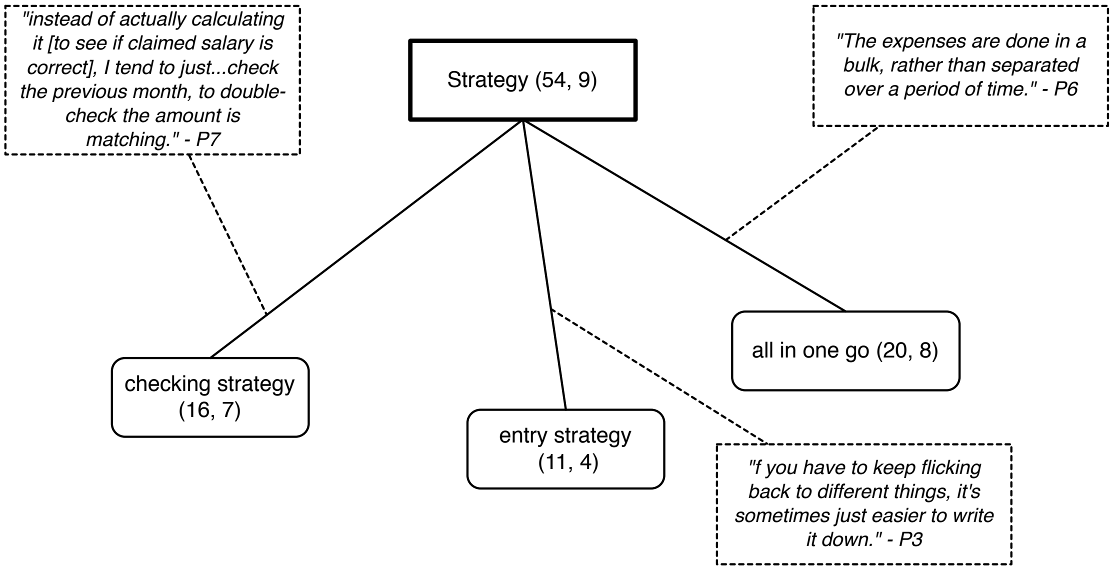


Figure 3.7: Diagram showing the theme Strategy.

Eight of nine interviewees saved up data to enter it all in one sequence. If they received forms with data input to check or enter, they saved these for later and then processed all the forms. P1 indicated he processed forms in batches of at least five forms, and found it disruptive to do just one or two and then switch to something else.

P2 was the only interviewee that processed forms with numbers to enter as they came in, but did admit that if she had more data to fill in, she would probably do it in a more efficient way.

Some people said they preferred batching their work so that they entered all data into a single system at once because it is quicker and easier to do the same type of tasks in one sequence, and when that is finished concentrate on another task. P4 mentioned he does them all at once because he gets the forms in a bulk and feels pressure by his boss to finish them all immediately, rather than spread them over time. P6 explained he postpones processing expense forms until the deadline to submit forms for that month has passed, after which he does all forms in one sequence.

People also talked about other strategies they used to do their job more efficiently. For instance, if they had to get out of the system to look up information digitally and then get back to the system window to enter it, they preferred to memorise the information, rather than flick back and forth and look it up each time they needed it. With numbers they had to enter frequently such as project codes, they memorised it even if they did not deliberately choose to do so. If the information to remember was complicated, they would write it down.

As discussed at the Checking section, people had different checking strategies and the number of times they looked back to the data source to check it against the data input varied. P7 said she sometimes deals with similar calculations, so she prefers to check the calculation she did last time rather than calculate it again.

People explained that a lot of numbers they enter are calculated from other numbers. Some people liked to write out and keep a record of their calculation, in case someone had any questions on how that number was calculated.

Participant	Quote/note
P3	*"I just try and do it in the quickest way...It's nice, once you've done it, it's completed, so it's sort your weight lifted [laughs]. So you don't need to think about it again."*
P6	*"the expenses are done in a bulk, rather than separated over a period of time. When I'm doing it lots at a time, I think once you get into sort of the hang of it, it gets done a lot quicker than..you just get used to putting them in, and inputting it all."*
P9	*"I try to concentrate on my task...I try to do one task [i.e. doing all expenses], finish one, and then do another."*
P4	It's difficult to take rests or even switch in-between number entry tasks because of the work pressure, and feels pressure by boss.

Table 3.11: Most participants entered all numbers in one go.

Participant	Quote/note
P3	*" I wouldn't necessarily have to [memorise numbers], It's more just if you have to keep flicking back to different things, it's sometimes just easier to write it down, or just try and remember it. But you can obviously take the long version and keep flicking back to the correct screen."*
P2	*"we have different grants and different project codes as a result, but you, because you use them so much, you end up remembering them."*

Table 3.12: Examples of strategies people used.

Importance of accuracy and paper trails

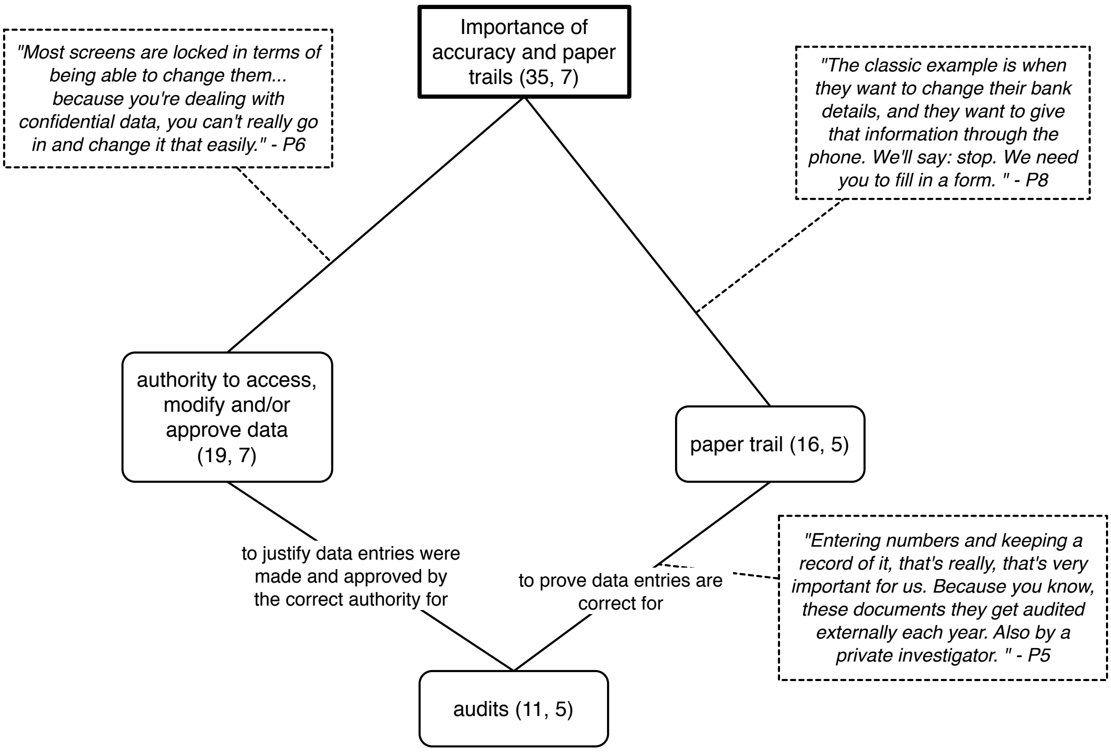


Figure 3.8: Diagram showing the theme 'Importance of accuracy and paper trails'.

Because interviewees were working with sensitive financial data, another theme was importance of accuracy and paper trails. Some data could only be entered in the system and modified by certain people. Large figures had to be approved by a superior first before it could be processed, so it was clear to an auditor that expenses were made with the correct approval. Hard copies of data had to be archived and were checked by external auditors. For instance, all expenses were claimed on paper forms, and had to have the original receipts as evidence that the expense claims were correct. Some data could not be given over the phone but had to be written down and sent via a paper form.

Other

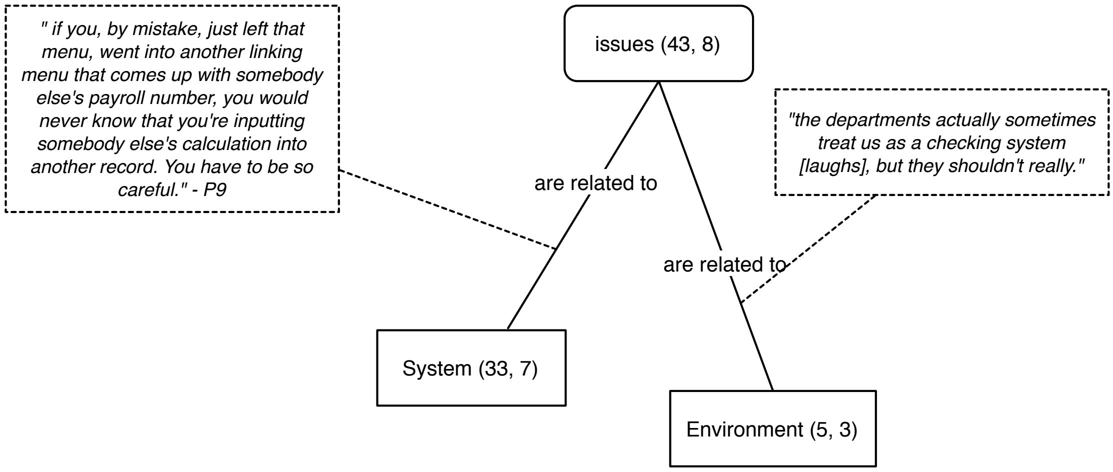


Figure 3.9: If people described issues, it usually had to do with the system.

Interviewees mentioned several issues they experienced. Most of the time, the issues had to do with the current data entry system they were using. P7 mentioned number entries on the computer screen were small and hard to read, and could not be enlarged. P5 mentioned she did not get notifications from the system if there was an issue with a payment, so she would not know about it until staff members would call up to complain they had not been paid yet. Other issues had to do with the organisation, such as the experience that the multiple checking layers made people rely on other people to check their input.

Participant	Quote/note
P5	*"There are other issues. You could say I think hundreds, I mean not just with the work that we do on expenses, but across [university A], across [university A] Finance, the Finance division...We just have to kind of work our way around the system and you know, adapt to it."*
P7	*"It's only the matter of how you get used to the Payroll system. Because companies have different systems, the data inputting can take a while to get used to it."*
P9	*"You know, all systems are a bit funny, I think. But you just gotta get used to it."*

Table 3.13: Issues that participants experienced with the system.

3.1.4 Discussion

The purpose of this study was to gain a better understanding of the type of data entry tasks in financial offices, and the physical environment in which these are conducted. For this purpose nine interviews were conducted with staff from two public universities who worked with financial data. The main findings of the study are:

- people have to get the data to enter from various sources, with varying information access costs
- people batch their work to enter a lot of data entry at once, and minimise switching between tasks.
- data entry errors are common, and the current solution is to have data entered and checked by multiple people before it gets submitted to the system

Information sources and access costs

A prevalent data entry task of participants' work was processing expense claims from university staff and students. A substantial part of this task is not just entering the data but also collecting it from multiple sources. For example, the expenses had to be entered from paper forms and receipts, foreign currencies had to be converted and copied from a website, and staff and financial information had to be retrieved from sources such as electronic spreadsheets, databases and e-mails. In data entry experiments, the data is often shown on a computer, to make it easy to manipulate. This was sometimes the case in the office setting studied here, but data was also transcribed from paper sheets into a computer. Furthermore, in lab experiments people are often only presented with the data they have to enter and sometimes are given only one data item at a time. The sources from which people had to enter data in this study usually contained a lot of data, not all of which was relevant to the task. The amount of irrelevant data on the sources can increase the time people need to look up the information they need.

The cost to access the information sources differed. For instance, some paper sheets were on people's desk, but some paper sheets had to be retrieved elsewhere. Sometimes participants had to go through a large spreadsheet, before they found the number they needed to copy. Participants also dealt with multiple windows on their computer screen, and sometimes needed to switch between different windows. Instead of flicking back and forth to view information they had to enter in another window, people said they preferred to memorise it. This supports previous research which showed that people make strategic use of internal and external resources and do not always maximise off-loading cognition to external resources (Gray et al., 2006). Even though participants were aware they did not have to remember information, they found this easier and faster rather than looking it up or writing it down.

This strategy allows people to be faster, but carries the risk that they misremember it. In previous studies, trying to hold more items in memory during a copying task increased errors (e.g. Morgan et al., 2009). However, in these studies people had to copy over coloured blocks, which were abstract to the participants and may have been too hard for participants to memorise (Waldron et al., 2011). In the current study, the information had some meaning to the users, and even if they were dealing with abstract information, such as project codes or ID numbers, some items were entered more often than others. Participants mentioned that even if they did not make a conscious effort to memorise these, they had to enter these numbers so often that they ended up remembering them. This is in line with prior findings that some numbers are more familiar than others, and that these familiar numbers are more strongly represented in memory (**?**).

If information to copy is familiar and contains numbers and text, making the effort to encode the information more deeply has been shown to make people more accurate (Gray & Fu, 2004; **?**). It is therefore not clear if the effect of a memory-based strategy on accuracy, as found in (Morgan et al., 2009), would extend to copying the material of this study.

People used some external tools to aid memory. For example, some participants had printed sheets with information they used and entered often, so they did not have to repeatedly look this up in the system. They also used calculators and pen and paper to perform calculations.

Batching data entry tasks

Eight of nine participants batched data entry tasks to do them all at once, rather than spread them over time. The only participant who processed expense forms as they came in said she did not receive a large amount of forms. She explained that if she dealt with a higher volume of expenses, she would probably do it in a more efficient way.

There were different reasons behind the strategy to enter a large amount of data in one sequence. One participant received forms in a bulk, and felt time pressure to finish them all at once, rather than spread them over time. The other seven participants said they received forms on an ad-hoc basis, but deliberately saved them until a specific moment and then processed them all in a bulk. They preferred to focus on one task at once, and some people stated that it made them faster in entering data after a while.

This is consistent with Healy et al.'s (2004) study, where learning effects could be measured in a number entry task. However, in their experiment participants became faster but also more erroneous in entering data. In the current study, P3 mentioned that most of the errors she picks up from colleagues are typos caused by rushing to finish the task. P4 mentioned that he wants to enter numbers accurately, but that there is a time pressure to finish things in time. This is in line with the speed-accuracy trade-off discussed in literature, and that a sense of urgency may cause people to enter numbers faster but at the expense of a higher error rate (Lin & Wu, 2011, 2013). Healy et al. (2004) recommended regular breaks when entering large amounts of data to maintain accuracy, but P4 in the current study explained it is difficult to initiate breaks himself, when he feels a time pressure to finish it.

Multiple checking stages

In order to prevent errors, input was entered and checked by multiple people before it was allowed to be submitted into the system. This checking method is similar to **?**'s (**?**) Swiss Cheese model, where multiple checking layers are used to minimise the risk of errors. However, as a result of this people seemed less careful in checking their own input, because they knew it would be checked by someone else.

In previous studies, providing incentives for people and introducing lockouts into the data entry interface has been shown to influence how carefully people check their input (Gould et al., 2015; Li et al., n.d.). The findings from the current study suggest that how and whether people check also partly depends on what the primary objective of their task is: entering or checking input.

When people were asked to check other people's data input, they checked each data item one by one. However, when they had to enter data themselves, they tended to chunk data items and only checked their entries when they had entered all items on one source. This finding is based on both observations when participants demonstrated a number entry task, as well as descriptions people gave on how they tend to perform their task. When people are asked to enter data, checking their input is not the main focus of the task, but rather a post-completion step that has to be performed after the main goal of entering data has been completed. It is well known post-completion steps are prone to errors, as it is not part of the main goal of the task (e.g. Byrne & Bovair, 1995; Li, Blandford, Cairns, & Young, 2008; Li et al., 2006).

It could also be that people were aware their input was going to be checked by another person, and therefore did not spend as much effort trying to check it properly. When talking about checking other people's input, people emphasised how it was important that each data item was checked carefully. However, when they talked about their own input, they tended to emphasise this less and mentioned another person would double-check it anyway.

Summary

The purpose of this study was to get a better understanding of data entry tasks associated with finance administration office work, and the context in which these tasks are done. While it was informed by previous data entry research, it did not have a particular theoretical framework prior to undertaking the study to inform and guide the data collection and analysis.

A prevalent data entry task of the participants was processing expense claims from university staff and students. A substantial part of this task was not just entering the data, but also collecting it from multiple sources with varying information access costs. As my data collection and analysis progressed, it emerged that framing the findings in terms of different information access costs (IAC) of these information sources had the potential to provide a useful lens on the data for informing the design and evaluation of data entry systems. As the study looked at the wider context of participants' data entry work and did not focus on entering expenses per se, it was not recorded which sources people needed and how they looked up information.

The aim of the next study is to investigate the information sources people need for an expenses task, and how they currently manage subtasks of looking up information from an expenses task.

3.2 Study 2: Information sources for entering expenses

3.2.1 Introduction

A large part of data entry work studied in Study 1 was not only entering the data, but retrieving it from the environment. People had to leave the data entry system to look up information from several sources, and hold data in memory whilst switching between documents, applications and the data entry system. The purpose of this study is to investigate people's current work practices to look up information from the environment for data entry work. What information do they need, and what information sources do they use? And how do they address these needs? Do they look up information as they need it, or get all the required information first and then enter it? It is important to understand how people manage subtasks of looking up information as part of data entry work, as different strategies impact task performance.

The following questions will be addressed in this study:

1. What is the information needed for an expenses task?
2. Where is this information retrieved from?
3. What are the strategies people use to look up information?

A contextual inquiry was conducted with nine finance employees from the same population as in Study 1. In this study, the specific focus was on people's information behaviour. Participants were observed at their workplace, and asked to first carry out a data entry task while thinking out loud. Next, they were observed while they continued working as they would normally.

Participants in the current study were video recorded while doing the expenses task. The video recordings captured the participants' interactions with the artefacts involved in the task, but the financial data on the information sources could not be identified from these recordings. The video recordings were used to supplement written observation notes, and after the observation part of the study, some video segments were played back to participants to explain what they were doing at certain moments.

Data was gathered and analysed using Distribution Cognition (DC) as a theoretical framework (Hutchins, 1995). Distributed Cognition is a theoretical framework that views cognition as distributed between people, internal and external sources and over time. As it takes the distributed nature of cognition as focus of analysis, it was considered it to be a useful framework for this study to help make sense of the fragmentation of, and access to, information for data entry work.

To analyse the data, I followed the guidelines of Furniss & Blandford (2006) on constructing descriptive models of the task environment. Though the models were originally developed to apply Distributed Cognition for teamwork, the models can also be useful with an individual as the focus of analysis. In particular the Physical and Artefact model were useful in understanding the distribution of, and access to, information for the user. The methodology facilitated to identify potential issues, and better understand people's current strategies and workarounds.

3.2.2 Method

Participants

Nine participants (five male) from four different finance offices took part in the study.

Participants from Study 1 were invited to participate again, but only one participant was able to participate again in the current study (P8 in Study 1, P1 in Study 2). For the remainder of recruitment, a parallel sampling approach was taken (**?**). This means participants were drawn from the same population using the same recruitment techniques, but were not the same individuals. As in Study 1, they were employees from financial offices at public universities dealing with processing expense claims. Participants had one of the following roles: research manager, payroll officer, and administrator.

People were recruited through a combination of convenience and snowball sampling. People were invited to participate via emails sent to opt-in mailing lists of Finance departments, and emails forwarded by contact persons and people who had already participated. A session with a single participant took between 2 and 2.5 hours, and participants were reimbursed with £15.

Site setting

The study took place in the same type of office setting as in Study 1. Participants were recruited from the same two public universities, and one office from Study 1 was visited again in this study. The other participants were from three different offices within these universities. They worked in open plan offices with two or more colleagues working nearby.

Contextual inquiry

My data collection was informed by using the methodological approach of contextual inquiry. Contextual inquiry is a combination of observation and interviewing users and their everyday work, with the aim to use the findings to inform design of the systems they use (Holtzblatt & Beyer, 2014). Holtzblatt & Beyer (2014) argue that observing participants carrying out their work can reveal concrete details, and it can help participants to recall past situations of carrying out work. It is therefore considered to be an appropriate method for the aims of this thesis, as it involves understanding users' expenses work with the aim to translate it into design implications for their expenses system.

All interviews were audio recorded, and observations were audio and video recorded. The contextual inquiry consisted of four parts Holtzblatt & Beyer (2014):

1. Interview

   In this part, the participant was asked general questions about their work. This part was primarily to make the participant feel comfortable and start the session easy. As I had already gained an insight on finance employees' general work in Study 1, this interview was shorter and more particularly focused on their expenses task and information sources they used.

2. Transition

   In this part, the participant demonstrated the expenses task while thinking out loud. I prompted the participant to elaborate what was happening if the participant fell quiet, or if something interesting or unusual happened.

3. Observation

   During this part, the participant continued doing his/her work without being interrupted or explaining what he/she was doing and why. I observed and made notes.

4. Summary

   In the final stage, I summarised to the participant what I had found to check if these assumptions were correct. If some parts of the observation need clarification, segments of the video recording were played back to the participant, and he/she was asked to explain what was happening at certain moments. The participant was thanked and debriefed.

Pilot study

In order to test the suitability of the study set-up, a pilot study was conducted with a financial administrator at a university who dealt with processing expenses. The study took place at the university, and notes were taken with pen and paper.

Initially, the intended method of this study was to conduct a contextual inquiry, followed by a week-long diary study where office workers would log diary entries of their expenses tasks. The aim of this diary study would be to get a further insight in additional information sources they may sometimes use, that were not covered in the contextual inquiry. Participants would submit a diary entry either by writing down a description or by taking a photograph of the task setting, showing the information sources. At the end of the day, they would have to answer the following questions about their short entries: what information did you need, where did you need to get it from, and when did you look up the information. This method built on a study by Sohn et al. (2008), where participants kept a diary of their mobile information needs and how they addressed those needs.

The administrator explained that expense tasks usually are conducted in the same manner, and predicted I was unlikely to find a lot of instances that differ from my observations of having office workers do the task. Furthermore, by observing them I am able to see the access people have to the resources and how much time it takes them to get the data, as well as when in the task they decide to look up information. This information would be more difficult to get insight to through diary entries.

It was therefore decided after this study to not conduct a diary study but instead only observe office workers, and ask them to explain their work.

Ethical considerations

The study has ethical approval from the UCL Research Ethics Committee [Project ID Number UCLIC/1415/001/Staff Brumby/Borghouts].

Participants were invited by email. The email included details such as the purpose, duration and reimbursement, and what participation would involve.

Before the start of each session, participants were again verbally briefed about the purpose of the study, and what would be expected of them. They were then asked to read and sign a consent form, and were given an information sheet with the study information and my contact details for them to keep.

All participants gave permission for the session to be audio and video recorded. After the video camera was set up, participants were shown what was captured by the camera, to confirm the camera was set up appropriately. This ensured the participant that no financial data was legible from the recordings.

It was explained that the purpose of the study was to get a better understanding of how they normally go about their work with an aim to improve the system, and that there were no right or wrong ways of doing tasks. They were free to withdraw from the study at any time. Participants were informed that the data would be used for research purposes only and stored in accordance with the Data Protection Act 1998. Names of participants and the organisations they were working at were not included in interview notes and transcripts.

Video recordings were not only used to supplement audio transcripts and written observation notes, but were also used in debriefing interviews. If something interesting or unusual happened during the observation, a video clip of this moment was played back to the user in the debriefing, and they were asked to elaborate on this moment.

The use of video recordings to aid recall in the debriefing interviews was inspired by **?**, who captured screen recordings to capture workers' activities in a law office. In debriefing interviews, they played these recordings back to the workers, and asked them to explain their activities. The recordings were useful for workers to accurately recall what they were doing at the captured moments, and why they had certain windows open.

Screen recordings provide a more detailed account of activity than video recordings, however there are also privacy issues and not all participants agree their activity to be recorded (Rule, Tabard, Boyd, & Hollan, 2015). Due to confidentiality issues to share financial data, it was not possible to use screen recordings in this setting.

Data collection and analysis

All sessions were audio and video recorded.

Every time the participant used an artefact, I asked them to show it to me. Artefacts included paper, a computer program, a tool such as a calculator, or a spreadsheet. I wrote down a quick description of the artefact and where possible, a photo or screenshot was taken of the artefact.

Screenshots were made of the data entry system participants used to enter the expenses data These screenshots did not include any data entries.

In addition to video recording the think-aloud and observations parts, I made notes by hand. Whenever something interesting happened during the observation part, I made a note of it to remind me to ask the participant questions on this in the debriefing interview.

Data was analysed using a Distributed Cognition (DC) perspective. I used Furniss & Blandford (2006)'s guidelines in developing the following descriptive DC models:

- The physical model: this model describes the physical layout of the task environment
- The information flow model: this model describes how information flows through all users involved in the task
- The artefact model: this model describes all artefacts involved in the task

These models are based on the working models of contextual design to identify work activities, but are more focused on how information is distributed in the environment. Each model consists of a diagrammatic representation that visualises the data, and a narrative representation which verbally describes the data.

After each data collection session, written notes were typed out and any initial thoughts or findings were added. After data collection from the first four participants, initial versions of the models were made. Making these initial versions helped identify any gaps in the data collected so far, and helped guide further data collection sessions.

Audio recordings of the interviews and think-aloud verbal protocols were transcribed verbatim. Video recordings were played back, and additional notes were made if anything new was observed by watching these video recordings.

The written transcripts and notes were read and if a quote or note was relevant to one of the models, it was grouped under that model. The groupings were reviewed and used to refine and expand the models. Video recordings were consulted to fill any gaps from the written data.

3.2.3 Findings

I used the DC principles as described by (Furniss & Blandford, 2006) as guidelines to decide what to include in the models. The principles of each model are described at the start of each section, and marked in italics (e.g., *horizon of observation*) in the narrative descriptions.

Physical model

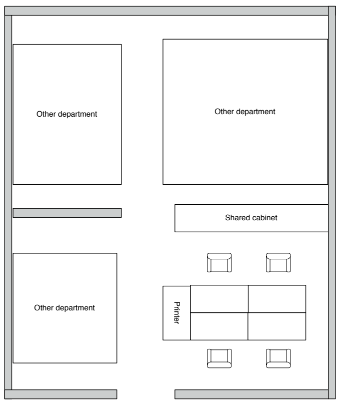


Figure 3.10: Physical model diagram showing a typical physical layout of people's work environment at room level.

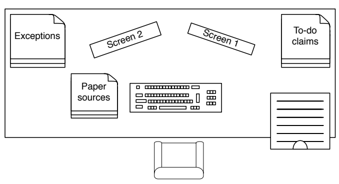


Figure 3.11: Physical model diagram showing the physical layout of people's work environment at desk level.

The physical model describes what the individual can physically hear, see, and access, and how information sources are laid out in the physical environment. In developing the model the following was considered: what is the proximity of, and access to, devices and people: what can be seen and heard from the individual's point of view?

Physical model principles

- Space and Cognition: how do people use the physical space to support their work
- Subtle bodily supports: do people use their body to support their work
- Situation awareness: are people informed of what is going on
- Horizon of observation: what are people able to see and hear
- Arrangement of equipment: how do people arrange their equipment

The physical layouts of the four offices were not identical, but shared a number of characteristics. All participants had their own desk and worked in an open office with two or more colleagues. P3 was the only person in the office responsible for data entry tasks. The other eight participants had colleagues in the same room who dealt with similar tasks. P1, P2, P8 and P9 also had colleagues in the same room that worked on different tasks. Colleagues that were working on different tasks were situated further away from them than other colleagues. As an example, the physical environment of the office of P1 and P2 is depicted in Figure 3.11 at room (a) and desk (b) level. The number of people in one room ranged from two to sixteen.

The open layout made it easy for colleagues to interact with each other and share information between themselves. They could see when a colleague was present and available to consult. Participants regularly consulted colleagues in their room to retrieve information they could not find on any other sources. This information is given verbally, or a colleague directs the participant to the correct information source.

The offices had an open-door policy, meaning that workers could at any time be interrupted by people walking in. During the observations, participants were regularly interrupted by colleagues. They responded to it but used *subtle bodily supports* to not lose track of where they were in the expenses task. For example, P3 responded verbally to a colleague but kept his kept his visual attention on the computer and tried to continue with the expenses task. When P1 was interrupted by a colleague, he placed a finger on the computer screen to remember where he was in the task.

All participants worked with both paper and digital information sources. Several participants used the physical *space* to organise their paper sources. P2 maintained separate physical locations on his desk for expense claims to be processed, expense claims that had been completed, and exceptional claims that required further attention. P7 also had paper sources fixed on his wall, which were visible from his desk.

Other physical sources were located in an individual's drawer, in a shared bookcase or drawer in the same physical room. These were not visible from their desk, and participants had to physically move from their desk to retrieve and view this information. Participants could see whether a colleague is present or not, and whether they can consult them at that moment to request information. Recent physical files are kept in a closet in the same room, whereas older files and employee files are physically kept in another location. These older files were used less frequently. If P6 received expense claims, she placed these in her drawer to return to later. If there were a lot of claims to be processed or if the participant started processing claims, these were placed on their desk.

Participants become aware of new claim requests by batches of claim forms or receipts on their desk with a handwritten note, the claimant asking them in person, or when they check their e-mail and see a new e-mail by the claimant. For physical claims, they browse through them and place them either at a dedicated area on their desk or in their drawer, to return to later. P2, P3 and P6 re-ordered claims based on urgency. Claims that were more urgent were processed first. These could be claims by their boss, or claims with an explicit instruction that it was urgent. In addition, P2 sometimes grouped claims according to category. The reason for grouping them is that different categories of expenses can require other data fields to be filled in, and P2 felt it to be easier to do these particular type of claims together. Examples are travel expenses for which participants have to fill in the departure point and destination, or external lunches for which the name of the restaurant has to be filled in. Participants place exceptional claims in a separate physical location. For P2, this was in a separate tray in the top-left hand corner of his desk. By putting it in his *horizon of observation*, he can see if there are still claims in this tray to be processed and return to at a later point in time. P2 starts each day by seeing if there are claims in this tray, and processes these first. P5 places exceptional claims in a separate drawer. They are not visible and she does not use any reminders, but returns to these when she sees fit.

Participants also received requests digitally through email. They could not place these in their horizon of observation. However, whether the claim requests were in physical or digital form, they always needed the physical receipts of expenses, before they could process a claim. Similar to physical claim forms, they had these receipts either in their drawer or on their desk, as a reminder that they still needed to process these claim requests.

Participants also created their own artefacts to aid them in their work. This includes physical artefacts, such as a spreadsheet with frequently used codes, as well as digital artefacts, such as a digital form to look up codes by person name. Participants could type in a name of a claimant, and the form populated codes related to one person. This made it easier for people to look up which account codes to charge the expenses to.

Digital information sources included the data entry system, intranet, and external websites via their computer. Furthermore, they had digital files, such as PDF files, Excel spreadsheets, and Word documents stored on their desktop computer. The *arrangement of equipment* was that seven out of nine participants had two screens. All seven participants mainly used one screen, both to enter and retrieve information. Some participants used a second screen to display their email inbox. If they received a notification of a new email, they would briefly glance to determine the urgency and importance of the email, but would try to continue with the expenses task.

Sometimes a claim was checked by multiple people in the same room. Participants had *situation awareness* of the progress of the claim as long as it was still in the same physical office. They could see the progress of expense claims by the pile on colleagues' desks, and could ask colleagues in person. Furthermore, they can overhear conversations on the phone and become aware of claims that require further attention. Participants can see their own screens and desks. They have to walk over or do additional physical actions in order to view what is on colleagues' desk.

As soon as an expense claim was submitted to another office and physical location, the user would have little *situation awareness* of the progress of the claim. The system did not have a visible status update, meaning participants can not see the progress of their claim once it has been submitted to Central Finance. As they do not receive updates, administrators and claimants often forget to keep paying attention to it and do not query it until they realise payment has not been processed yet. Often an error will have occurred and at this point the project may have finished and payment is no longer possible. If participants needed more information on the status of a claim, they needed to contact colleagues from another office via email and telephone, or they had to visit the other physical location. Participants called colleagues from other departments with queries such as errors, and outstanding claims that had not been processed yet. They emailed claimants with queries such as if they do not agree with the expenses claimed, if they need further information, or if they have spotted an error.

Artefact model

The artefact model describes the artefacts that are used.

Artefact model principles

- Mediating artefacts: do people use any artefacts to support their work
- Creating scaffolding: do people use the environment to simplify tasks, e.g. set reminders
- Representation:goal parity: do people use artefacts to display the explicit relationship between the current and goal state of their work
- Coordination of resources: how do people coordinate their information

Participants work with both physical and digital artefacts. Table **??** shows an overview of the information sources that were involved in an expenses task. For each instance of the task, several artefacts can be used, but there are two main artefacts that are used in each instance and are central to the expenses task: the paper receipts and the expenses entry system. In addition to information sources, participants use several *mediating artefacts* to support their work. Calculators are used to aid in calculating sums. Multiple tools were consulted to convert currencies: an external website, a tool on the intranet, and a tool included in the data-entry system itself. They use a physical tray on their desk to hold exceptional claims. They use a pen to annotate receipts and highlight which items on receipts to claim back.

Some of the participants *created external scaffolding* to simplify their task. In particular, participants had difficulties remembering codes they needed to enter. P5 and P6 had made a personal spreadsheet with codes they used most frequently. P4 remembered old codes but had difficulties remembering new codes since they had changed 18 months ago. To look up codes, she used a spreadsheet created by the departmental manager where she could fill in old codes, that would populate the correct new code.

The parity between the current and goal state was displayed on the expenses system. Once a claim had been submitted, there would be a status update in the expenses system. This status reveals whether a claim is Pending, has been Paid, or whether Original receipts are required. As long as a claim was Pending however, there was however no insight into what is happening on the other side. Often a claim would be received but would be held because there was information missing. Participants would not know about this unless they explicitly contacted the office, and depending on the situation, they would often still not receive information on what was happening with a particular claim because the information was not centralised in Central Office and the person on the phone would not be able to look up what was happening with a particular claim.

In order to *coordinate resources*, the intranet was intended as a central point for claimants, administrators and Central Finance officers to access the same information resources. Participants find the resources difficult to use, so people often end up making their own local copies and working with these instead. This supports them better in their work, but they can end up working with old and incorrect information if information gets updated. Examples of information sources used locally are claim forms: claimants keep working with local copies on their computer and do not download the new forms. They keep using old information such as old salaries and old project codes. Another example are spreadsheets with budget codes: administrators created and used their own spreadsheets with codes. If codes were updated, they were not aware and ended up working with old codes. There was an instruction manual to do expenses, to ensure everyone carried this routine task in the same way. Newcomers usually learnt how to do it from other people rather than this written manual, as the experience was that it was often easier and faster to learn it this way. This way of learning the activity again had the risk that some people were doing it in the old, incorrect way, and passed on this incorrect way of doing work to others. In order to provide proof of expenses, people still had to provide hard-copy receipts. These need to be sent to Central Finance, but often get damaged and lost and the claimant and administrator will not know, but Central Finance will not know either so nothing will happen unless the administrator or claimant takes action and chases it.

In order to prevent people from interrupting an expenses task, people were logged out of the data entry system after a period of inactive use and they had to restart the task from the beginning. This added cost to resume the task kept participants focused on the data entry task, and they were less likely to interrupt and switch to unrelated tasks. However, people often did not know beforehand what the cost to access information was going to be. Furthermore, it was also not clear after how long the timeout would occur [quote].

Information flow model

This model describes the flow of information through several actors for an expenses task.

Information flow principles

- Information movement: how does information move throughout the system
- Information transformation: does the representation of information undergo changes
- Information hubs: what are the main points where different information channels meet
- Buffering: are there any buffers to uphold information
- Communication bandwidth: how does communication take place
- Informal communication: does informal communication take place in addition to formal communication
- Behavioural trigger factors: are there any local factors that individuals respond to

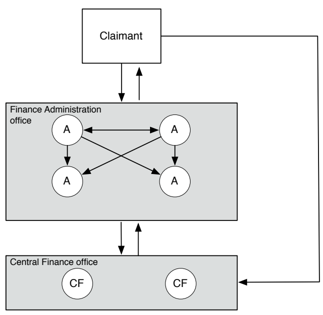


Figure 3.12: Model of the information flow.

An expenses claim *moves* through several actors, and moves between actors via email, phone, physical post, and face-to-face communication. The actors who contribute to the processing of an expense claim have limited visibility on the overall status and progress of the claim. For example, once claimants submit a claim request to the administrator, they do not know what the status is of that claim until they receive an email notification that it has been completed. Similarly, once administrators submit a claim to the Central Finance office, they do not know what the status is of the claim. They do not know when or whether it has been processed and if not, what the reasons are for holding it. The workers at the Central Finance office know the reasons for withholding a claim, but often receive incomplete information of a claim, and for example do not know the justification behind a claim, whether the expenses are made correctly and if there is an error in the project code entry, they do not know the correct project to charge it to.

Information transformation takes place when calculations have to be carried out. At the beginning the individual numbers are saved, as well as the calculations on those numbers. Once a claim is submitted only the end result will be saved on the system. For example, if one claim request involves multiple expenses, each individual amount has to be checked by the administrator. Administrators are then free to choose whether to type each amount or only the sum total on the system. Once they have processed and submitted the claim, only the sum amount will be available for auditors.

There are two main *information hubs*: the first hub is the office of the administrator, who deals with incoming claims and hard-copy receipts from claimants. These claims are processed at the administrator office and then sent off to the Central Finance office, the second information hub. Workers at this office deal with incoming claims and hard-copy receipts from administrators, match and process these and submit them for payment.

The administrator is the main *information buffer* between the claimant and the Central Finance officers. Claimants submit a request to administrators. This claim is upheld until the administrator decides to process it and send it to Central Finance. If there was an issue with a claim, claimants contacted the administrator, who then contacted Central Finance. Though claimants could also contact Central Finance directly, administrators said it was often easier if they contacted Central Finance on their behalf, as they knew who to contact [quote].

Communication between the claimant and administrator takes place face-to-face, over the phone, via email and via handwritten notes. Communication between colleagues takes place face-to-face. Communication between the administrator and Central Finance solely takes part via email or over the phone, though it is possible to take place face-to-face.

Instructions are mostly *communicated informally* through word of mouth. Knowledge of how to use the system sits with the employees, and it is often faster to explain newcomers how to do it rather than go through the written instructions. A consequence is that when information gets updated, not everyone is aware of it and keeps using the old and incorrect way, or learn the incorrect way from someone else who is still using the old way.

Receiving a claim request from another actor are the main *factors triggering behaviour.* Participants collected claim requests and saved them to return to later. Some participants kept claims to be completed on their desk. The size of this pile acted as a trigger to decide whether to start processing them. Furthermore, the payroll deadline was another trigger. Participants tried to complete claims before the deadline so claimants were reimbursed in time.

Task strategies

By describing the task environment, several strategies were uncovered. These are discussed in more detail below. After developing the models, the audio transcripts, notes, and video recordings were reviewed again to identify patterns of information strategies.

Planning for information needs Participants started processing an expense claim by collected all physical sources they knew they were going to need. These were the paper receipts, and could include a physical claim form and notes with written instructions. They placed this on a pile on their desk next to their computer, and entered this information first.

Participants did not collect digital sources beforehand. Instead, participants retrieved these sources at the moment in the task they realised they needed it. When they realised they needed information from a digital source, they interrupted entering data, left the data entry window and went to search for the information.

Participants did not always know beforehand what the cost of collecting this data was going to be, but assumed it could be retrieved fairly quickly. Sometimes it could take longer than expected due to various reasons. First, they did not always know which source to consult for finding the data. For example, people in Central Finance had to validate if the person signing off a claim form was authorised to give this signatory. The information to check this could be in a spreadsheet, but was sometimes also in a different PDF file. At other times, this information had to be looked up on the departmental intranet. Second, if participants did know which specific source to access, they did not always know the associated cost to access it. If information about a certain employee needed to be looked up, people consulted a search engine and typed in the person?s name. Sometimes they found the information quickly, but sometimes it took a while before they found what they needed. Third, even if participants did know the specific source and the normal cost, this cost was not always the same. For example, a website could take longer to load than usual.

If information was difficult to find, participants had time thresholds to decide whether to postpone it and come back to it later. P2 said that if he felt he was spending more than ten minutes on a task, he would postpone it. He placed uncompleted expense claims that required further attention in a separate tray on his desk. He revisited this pile the next morning and tried to process them then. During observations, P5 could not find certain information for a claim either. After approximately five minutes of trying to find the information, she decided to write an email to the claimant requesting the information. She then put the claim form aside and started the next expenses task.

Creation of information sources Participants had access to shared files on the intranet of the office. They did not always find this information easy to use, and sometimes made their own local copies.

Deferral of interruptions In order to prevent people from interrupting an expenses task, people were logged out of the data entry system after a period of inactive use and they had to restart the task from the beginning. This added cost to resume the task kept participants focused on the data entry task, and they were less likely to interrupt and switch to unrelated tasks. Participants did however interrupt a data entry task for task-related purposes, such as looking up information. They left the data entry interface and opened a new window. As windows were maximised, they were unable to see the data entry interface whilst they were looking up information. Participants explained it was much easier and faster to look up information on the screen they were already interacting with, rather than switch screens. They also felt that, opposed to physical sources, they could retrieve information from digital sources quickly. It sometimes took longer than expected however to look up certain information. Furthermore, it was also not clear how long the system would wait before logging them out. Upon coming back to the data entry interface, participants often found that they were logged out unexpectedly, needed to log back in and start from the beginning. In most cases, their information was lost.

Use of multiple screens People dedicated one screen for the expenses task and maximised their window, so it filled the entire screen. This is in accordance with Bi and Balakrishnan (2009), who found that when dealing with two screens, people dedicate one computer screen to the primary task. However, they found that they use a second screen for subtasks. This was also found by Dearman and Pierce (2008)'s study on how people use multiple devices found that people assign sub-tasks to secondary devices to minimise the need to transfer information between devices. People in the current study often also used their primary screen to look up information for the expenses task, and switched back and forth between maximised windows, rather than look up and display information on the second screen. In contrast, previous studies on the use of multiple screens showed that people dedicated a second screen to look up information (Bi & Balakrishnan, 2009; **?**). Even if people knew beforehand which digital information they were going to need, they often started the task and looked up information as they needed it. For paper information, they did collect all information they knew they were going to need, such as the paper receipts, the claim form, and any additional post-its with instructions. Whilst there was a possibility to place information on the second screen, this is also time-consuming (Bardram et al., 2006). With paper sources, it is perhaps less time-consuming: no time was spent on arranging the sources on the physical space of the desk, but they were stacked in a pile on their desk or lap and the right source was picked out when needed. Dearman & Pierce (2008)'s study on how people use multiple devices found that people assign sub-tasks to secondary devices to minimise the need to transfer information between devices. Potentially people used the primary screen for looking up information so the information was on one screen and try could try to copy and paste it. Often though this was either not possible or users chose themselves to manually transcribe it.

People used additional screens for other tasks. For example, the second screen was often used to display the email inbox, but this was not consulted during the expenses task. Even in instances when people had to look up information from an email, they would open their inbox on the primary screen, rather than look it up on the second screen.

3.2.4 Discussion

The purpose of this study was to investigate the information sources people need for an expenses task, and how they currently manage subtasks of looking up information from an expenses task. Do they look up information as they need it, or get all the required information first and then enter it? Do they change their strategies as they get more experienced with the task and know where to get the information from?

The following questions were addressed:

1. What is the information needed for an expenses task?
2. Where is this information retrieved from?
3. What are the strategies people use to look up information?

Information sources

The information needed for an expenses task needed to be retrieved from both paper sources, such as paper receipts, handwritten instructions and claim forms, as well as digital sources, such as spreadsheets, websites and emails.

The user did not always know beforehand what the cost of accessing digital sources was going to be. First, they did not always know which source to consult for finding the data. For example, people in Central Finance had to validate if the person signing off a claim form was authorised to give this signatory. The information to check this could be in a spreadsheet, but was sometimes also in a different PDF file. At other times, this information had to be looked up on the departmental intranet. Second, if participants did know which specific source to access, they did not always know the associated cost to access it. If information about a certain employee needed to be looked up, people consulted a search engine and typed in the person?s name. Sometimes they found the information quickly, but sometimes it took a while before they found what they needed. Third, even if participants did know the specific source and the normal cost, this cost was not always the same. For example, a website could take longer to load than usual.

Information was centrally available, but this was perceived as being difficult to use. As a result, users made their own local copies of information they needed and used these sources instead. Furthermore, procedural information was passed on via colleagues rather than the central information sources. This caused people to use outdated information. Multiple departments were involved in the task. There was no transparency of information and progress of activity, and explicit information exchange was needed. The system had a timeout to prevent long interruptions, but people still looked up information as they needed it. They often did not know the associated IAC and as a result were locked out.

Information strategies

With paper sources, no time was spent on arranging the sources on the physical space of the desk, but they were stacked in a pile on their desk or lap and the right source was picked out when needed.

People did not always know which information they were going to need for a task, which made it challenging to collect all information beforehand. Therefore, people usually started a task collecting physical sources they knew were needed, and looked up other information as they needed it. Even if people knew beforehand which digital sources they were going to need, they still looked up information as they needed it.

This is in contrast with Sohn et al. (2008), who found that uncertainty of the location of information would cause people to leave looking for it until later. A difference with this study is that Sohn et al. (2008) looked at people?s information search behaviour for personal tasks. Perhaps there was a pressure for employees to finish their work tasks and the urgency of finishing the task weighed more than the potential time cost of looking up information.

People dedicated one screen for the expenses task and maximised their window, so it filled the entire screen. People in the current study often also used their primary screen to look up information for the expenses task, and switched back and forth between maximised windows, rather than look up and display information on the second screen.

Previous studies on the use of multiple screens showed that people dedicated a second screen for subtasks that supported the main task, such as looking up information (Bi & Balakrishnan, 2009; Dearman & Pierce, 2008; Grudin, 2001). Even if people knew beforehand which digital sources they were going to need, they often started the task without placing this on the second screen. They did not look up information until they needed it, and used their main screen for this.

Whilst there was a possibility to place information on the second screen, this is also time-consuming (Bardram et al., 2006). **?**'s study on how people use multiple devices found that people assign sub-tasks to secondary devices to minimise the need to transfer information between devices. Potentially people used the primary screen for looking up information so the information was on one screen and try could try to copy and paste it. Often though this was either not possible or users chose themselves to manually transcribe it.

People used additional screens for other tasks. For example, the second screen was often used to display the email inbox, but this was not consulted during the expenses task. Even in instances when people had to look up information from an email, they would open their inbox on the primary screen, rather than look it up on the second screen. This contrasts (Grudin, 2001) where people deliberately used a second screen to look up information, even if it was faster to retrieve it from their primary screen.

Other issues

Contribution

- Evidence that an expenses task is fragmented: people often have to go in and out of the expenses system to look up information.
- Evidence that the IAC of the required information sources for an expenses task varies.
- Evidence that depending on people's experience and awareness of how costly it is to access information, they will try to minimise switching between tasks.

The goal of this study was to give an understanding of how people manage subtasks of looking up information as part of data entry work in a financial office.

Limitations

The goal of this study was to get an understanding of how people collect information as part of data entry work in a financial office. By using a Distributed Cognition approach, it became clear that information for an expenses task was not only distributed between external artefacts and internal cognition of one individual, but also between people. As the focus of analysis of this study was on the individual and not the team, the human agents are included as additional information sources. Studying distribution of cognition from a teamwork point of view is beyond the scope of this thesis, but will be useful to study in future work. Several issues became clear that related to teamwork, such as limitations in communication and coordination of central information.

Due to confidentiality issues surrounding financial data, it was not possible to install logging software on participants' computers and all observational data reported here is qualitative. Given the situated nature of the study, it is also not clear to what extent people's behaviour is shaped by the access of the information sources, and how they are influenced by other situational factors such as user expertise. The next series of studies are set in a controlled environment, with the aim to study the influence of access costs to information on people's strategies, and how strategies impact quantitative performance measures as accuracy and speed. Study 1 and 2 have given an understanding of the task context and the information sources. The materials used in the subsequent studies will be designed to look similar to the office setting.

Chapter 4

Understanding the effect of IAC on looking up information

Chapter outline

The previous chapter demonstrated that for an expenses task, people have to copy data from multiple sources that differ in terms of their IAC. This chapter describes two studies that explore the extent to which IACs influence strategies, speed and accuracy.

Study 3 tests how IAC affects switching behaviour when copying from one source. It tested whether the effect of IAC on copying colours, as shown in previous studies, extends to copying numbers. Study 4 tests how IAC affects switching behaviour when information for one entry task has to be collected from multiple sources. Together these studies intend to show that IAC affects how people switch between entering and looking up information.

4.1 Study 3: Copying numbers from one source

4.1.1 Introduction

Study 1 and 2 illustrated that people had to retrieve and copy data from multiple sources with varying information access costs (IAC). People were not always able to look at the data source while entering data, and information sometimes had to be briefly kept in memory before people could enter it. In this situation, people have various options: they can decide to keep going back to the target source, reduce their memory load by writing it down, or keep as much of it in memory. Participants in Study 1 explained they tried to do it in the quickest way possible, and preferred to keep data in memory as this was often a faster strategy than writing it down, printing it out, or going back and forth between the source and the input window.

This is in line with previous lab studies, which showed that as the cost to access information increases, people increasingly rely on the information in their head rather than the information in the environment (e.g. Gray et al., 2006; Morgan et al., 2009). People look at the target source longer before copying anything, but after one look, they do enter more items in one sequence before looking back.

Gray & Fu (2004) conducted a study where people had to copy over VCR programming information. Participants either had permanent access to the information, the information was covered by a grey box which could be uncovered by hovering over it with the cursor, or they were explicitly instructed and trained before the trial to memorise the information. The people in the latter condition were more accurate than the other conditions in entering the information. This shows that a deeper encoding of the information in memory makes people more accurate in entering the information.

In later studies that further explored the effect of access costs on use of memory, the Blocks World Task (BWT) was used as a task paradigm, which requires people to copy a 3x3 grid of coloured blocks. In these studies, participants relied more on memory as the cost to access the target window increased. In one study, this memory-based strategy made participants better able to resume after interruptions by copying more blocks before having to revisit the target window (Morgan et al., 2009). Looking at overall task performance overall however, they did make more errors overall and took considerably longer to complete the task. In a later paper, the researchers reflected that the coloured blocks participants had to copy may have been too demanding to memorise (Waldron et al., 2011). This abstract visuo-spatial information did not bear any meaning to the participant, in contrast with the VCR programming information used in Gray & Fu (2004) which is more familiar and easy to memorise. This suggests that the type of information to copy influences the effect that IAC might have on task performance, though the studies differed in task paradigm making it hard to compare their findings: in Gray & Fu (2004) people were explicitly instructed to memorise the information and conducted a test prior to a trial during which they had to fill in the information, and could not continue until they had stated everything correctly. This ensured that people had the information well-memorised before they started the experimental trial. In the Blocks World Task studies, participants were not given this training or instruction.

Rather than training participants to memorise the information, **?** used an alternative approach to motivate people to encode the information more deeply. They conducted two studies where people had to transcribe text and numbers that were presented either in a black font colour or a harder-to-read grey font colour. Participants made fewer data transcription errors if data was shown in the harder-to-read font colour, both for transcribing text and numbers. In line with Gray & Fu (2004), these studies showed that when people do make the effort to more deeply encode information in a copying task, it can improve their accuracy.

In order to design data entry interfaces that support how people enter expenses from information sources with varying IACs, it is important to understand how these differences in IAC affect strategy and task performance. Relying on information in the head over information in the world can make it more likely for people to make errors in a copying task(Morgan et al., 2009), though other studies have shown that a deeper encoding of information can also improve accuracy if memorised well (Gray & Fu, 2004; **?**).

The study reported in this chapter replicates the BWT study, by using both coloured and numbered blocks.The purpose of this study is to see whether the effect on IAC on people's strategies and performance, as reported in previous BWT studies, holds when people have to copy numbers instead of colours.

It is expected that the effect of IAC on strategy will be similar and that an increase in IAC will encourage people to adopt a more memory-based strategy. However, the expectation is that numbers are easier to memorise and people will be able to memorise more items. Furthermore, based on previous findings that a deeper encoding of numbers in memory can reduce errors, it is expected that an increased IAC will improve accuracy for copying numbers.

4.1.2 Method

Participants

Fourty-two participants (eight male) were recruited from the UCL Psychology Subject Pool. Ages ranged from 18 to 52 with a mean age of 22.38 (SD = 7.45). Participants received course credit or £3.75 as a compensation for taking part in the study.

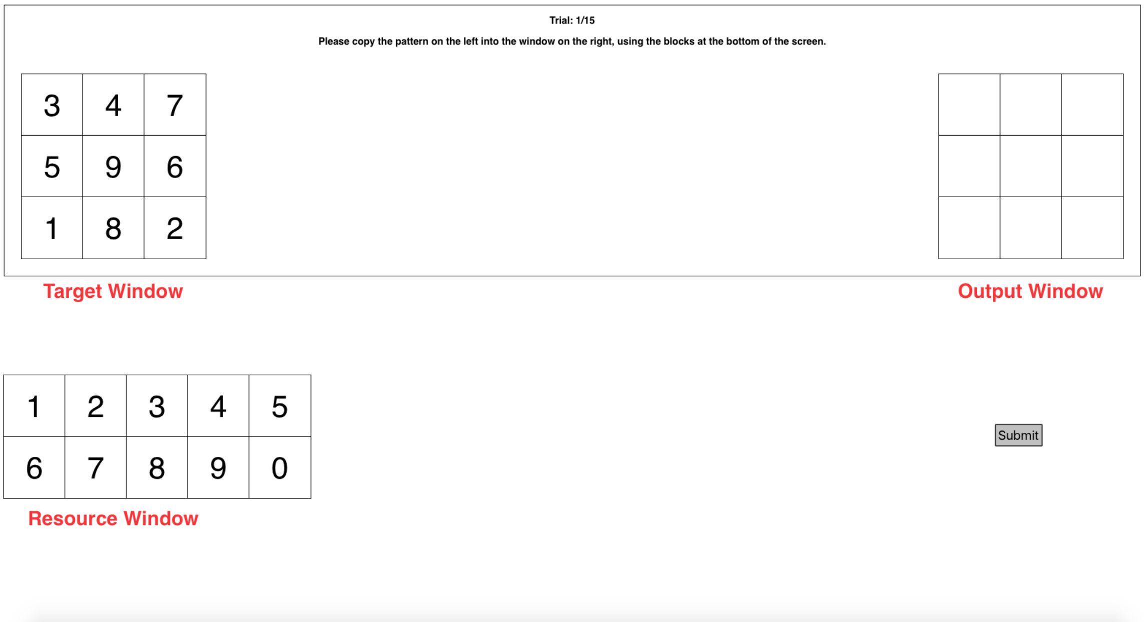
Design

A mixed design was used with two independent variables: IAC and block type. The between-participants variable was the level of IAC which had three levels. If the IAC was Low, the target pattern was permanently visible. In the Medium and High IAC conditions, the target pattern was covered with a grey mask, and could only be uncovered by moving the mouse cursor over the window. The mask reappeared as soon as the cursor left the window. In the High IAC condition, there was an additional 1-second delay to uncover the mask. This delay time was used in previous BWT studies where it showed to have a significant effect on task strategies and performance (Gray et al., 2006; Morgan et al., 2009; Waldron et al., 2007). The within-participants variable was the block type to be copied, which was either coloured or numbered blocks. The order was counter-balanced across participants.

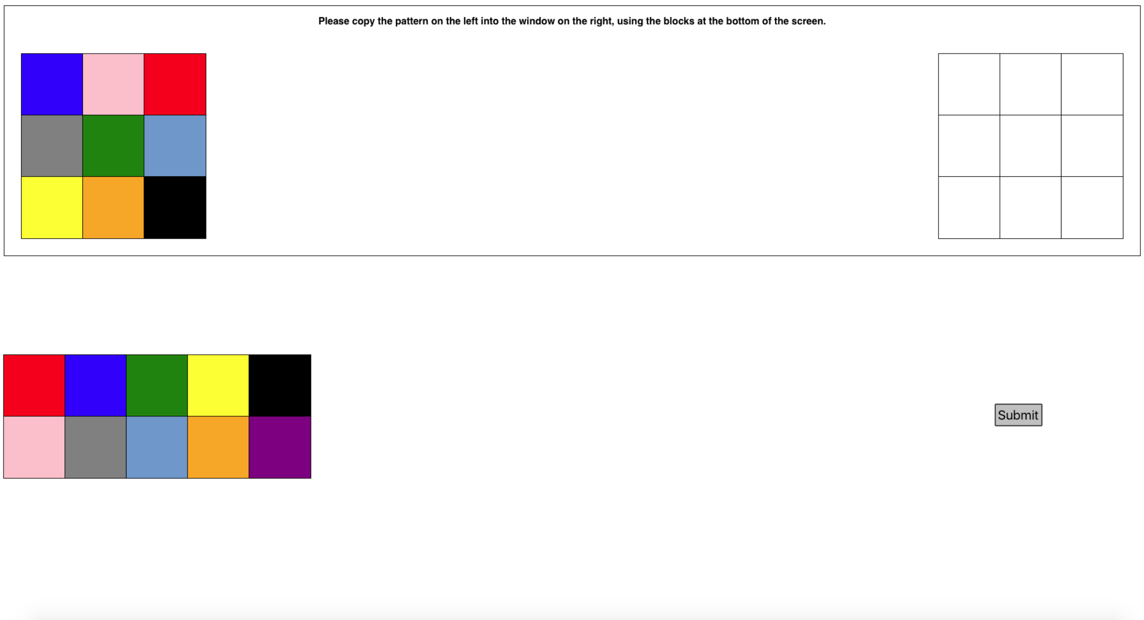
The dependent variables are listed in Table 4.1. The primary focus is on the measures of the first visit, as participants do not have any information yet on the target pattern. On subsequent visits, they may already have partial information in their head from previous visits. Therefore, the items copied after the first visit is believed to be the most 'sensitive measure of performance' (Janssen & Gray, 2012). Two dependent variables were used to measure accuracy. Incorrectly placed blocks measured instances where a participant initially placed a block in the incorrect place, but then moved this to the correct place prior to submitting the pattern. Incorrectly submitted trials measured instances where the participant had finished copying a pattern and clicked the Submit button, but the pattern was incorrect.

Strategy measures
Number of visits to target window
Visit duration of first visit (s)
Average duration of visits (s)
Number of blocks copied after first visit
Number of blocks copied correctly after first visit
Global task performance measures
Number of incorrectly placed blocks (per trial)
Number of incorrectly submitted trials (per experiment block)
Trial completion time incl. and excl. lockout (s)

Table 4.1: Dependent variables used in the study.



(a) The number condition.



(b) The colour condition.

Figure 4.1: The task lay-out with the three different components.

Materials

Figure 4.1 shows the task paradigm that was used. Each colour or number was only used once. The colours used were similar to the colours used in previous BWT studies (e.g. Gray et al., 2006; Morgan et al., 2009). Participants had to copy and complete fifteen patterns of each block type, and each participant had to copy over the same patterns. The target window showed a 3x3 grid with either coloured or numbered blocks. The output window showed an empty 3x3 grid, and was the same size as the target window. Participants had to copy the pattern shown in the target window by dragging blocks from the resource window and moving them into the output window.

The study was conducted on a desktop computer, using a 24-inch monitor with a resolution of 2048 x 1152 pixels. Participants used a computer mouse to drag and drop blocks. The experimental task was implemented using HTML, Javascript and PHP and run in a browser. All relevant browser events, such as mouse movements to (un)cover the grey mask, dragging and dropping the blocks and mouse clicks, were recorded and saved in a mySQL database. The browser window covered the whole screen to minimise distractions.

For the Low IAC condition, eye fixations were used to measure the number and duration of visits to the target window. Eyetracking data was also obtained for the Medium and High IAC conditions. However, this data was not used due to the fact that people were able to also view the target window area whilst the target pattern was covered. Therefore, in accordance with previous IAC studies (e.g. Gray & Fu, 2004), for the Medium and High IAC conditions the number and duration of uncovering the mask was taken as a measurement for visits to the target window. These uncoverings were measured by Javascript. The usefulness and limitations of using these measures are discussed in the Discussion.

A Tobii T60 eyetracker was used for recording people's eye fixations. Eye movements were recorded at a rate of 60 gaze data points per second for each eye, with an accuracy of 0.5 degrees and timestamp accuracy of 4 ms. For the analysis, all consecutive eye fixations with no drag or drop actions in-between were added together and counted as one fixation.

Procedure

Participants were welcomed and briefed about the experiment. It was explained they would be shown nine blocks which were in a certain order, and had to copy this order by moving blocks around. Participants were instructed to complete the task as fast as possible, but it was explained that they were not able to continue until they had copied a pattern correctly. The experiment was broken down in two parts, one where they had to copy colours, and one where they had to copy numbers. For each part, they were given two practice trials first to get familiar with the set-up, and to give them a chance to ask questions if anything was unclear. There was an opportunity for the participant to take a break between the two parts. They were then asked to read and sign a consent form and given an information sheet with a summary of the study and the researcher's contact details. In addition to the verbal briefing, the explanation of the study was written out on the computer screen for the participant to read and they were shown an instruction video that showed how the experiment worked. The study took around 20-30 minutes to complete.

Ethical considerations

The study was undertaken with ethical approval from the UCL Research Ethics Committee [Project ID Number UCLIC/1415/001/Staff Brumby/Borghouts]. At the start of each study, participants were first briefed verbally about the study. They were asked to read and sign a consent form, and were given an information sheet to keep. This information sheet contained a summary of the study information and the researchers' contact details. It was explained that an eyetracker would record their eye fixations and movements, but that these recordings were anonymous and that they would not be directly identifiable. After participants had completed the first part of the experiment, a prompt appeared on the screen advising them to take a short break. Participants could take a break as long as they wanted and could decide themselves when to continue with the second part of the experiment.

Participants were informed that the data would be used for research purposes only and stored in accordance with the Data Protection Act 1998. They were also informed that their data would be anonymised and when used in a report or academic paper, their data would not be directly identifiable.

4.1.3 Results

The means and standard deviations of all dependent variables are shown in Table 4.1.3.

Eight participants were removed from the analysis due to weak eye-tracking calibration. Furthermore, one participant misunderstood the experiment and did not know she was allowed to uncover the mask of the target window more than once. This participant had scores that were more than three times the interquartile range from the rest of the participants' scores on six different variables, so this participant was considered an outlier and removed from the analysis.

	Colours			**Numbers**		
	Low	Medium	High	Low	Medium	High
Strategy measures						
Number of visits to target window	**6.36**	**4.24**	**2.98**	**5.10**	**2.03**	**2.05**
	2.28	1.62	0.90	2.48	0.63	0.67
Visit time of first visit (s)	**0.39**	**0.04**	**2.18**	**0.51**	**0.04**	**1.49**
	0.23	0.02	1.59	0.45	0.05	1.01
Average time of visits (s)	**0.29**	**0.04**	**1.54**	**0.35**	**0.04**	**1.07**
	0.13	0.02	0.95	0.15	0.03	0.77
Number of blocks copied	**1.90**	**3.55**	**4.52**	**2.44**	**6.18**	**6.33**
after first visit	1.84	1.93	1.43	1.89	1.61	1.67
Number of blocks copied correctly	**1.86**	**3.22**	**4.07**	**2.36**	**5.96**	**5.98**
after first visit	1.75	1.83	1.20	1.74	1.52	1.55
Global task performance measures						
Number of incorrectly placed blocks	**0.15**	**0.67**	**0.79**	**0.17**	**0.31**	**0.46**
(per trial)	0.18	0.40	0.44	0.19	0.18	0.16
Number of incorrectly submitted	**0.27**	**1.9**	**2**	**0.36**	**0.5**	**0.83**
trials (per experiment block)	0.65	2.51	2.13	1.21	1.08	1.03
Trial completion time incl. lockout	**19.60**	**25.40**	**31.80**	**19.47**	**20.83**	**25.95**
(s)	2.98	5.16	6.08	3.03	3.08	4.21
Trial completion time excl. lockout	**19.60**	**25.40**	**28.84**	**19.47**	**20.83**	**23.89**
(s)	2.98	5.16	6.34	3.03	3.08	4.06

Task strategies

Number of visits to the target window

Participants made fewer visits to the target source when they had to copy numbers (M = 3.06, SD = 2.08) than when they had to copy colours (M = 4.49, SD = 2.18), F(1,30) = 41.62, p<.001, η^2 = 0.58. Participants also made fewer visits as IAC increased from Low (M = 5.73, SD = 2.41), to Med (M = 3.13, SD = 1.65), to High (M = 2.51, SD = 0.91), F(2,30) = 15.16, p<0.001, η^2 = 0.50. To investigate differences between conditions, post-hoc Tukey comparisons were performed. Results showed that participants made significantly fewer visits in the Medium-IAC condition than in the Low-IAC condition, p <.01. However, there was no difference in number of visits between the Medium-IAC and the High-IAC conditions, p=.59. Participants looked at the target window for colours more on every level of IAC (see Figure 4.2), and so there was no significant interaction, F(2,30) = 2.82, p=.08, η^2 = 0.16.

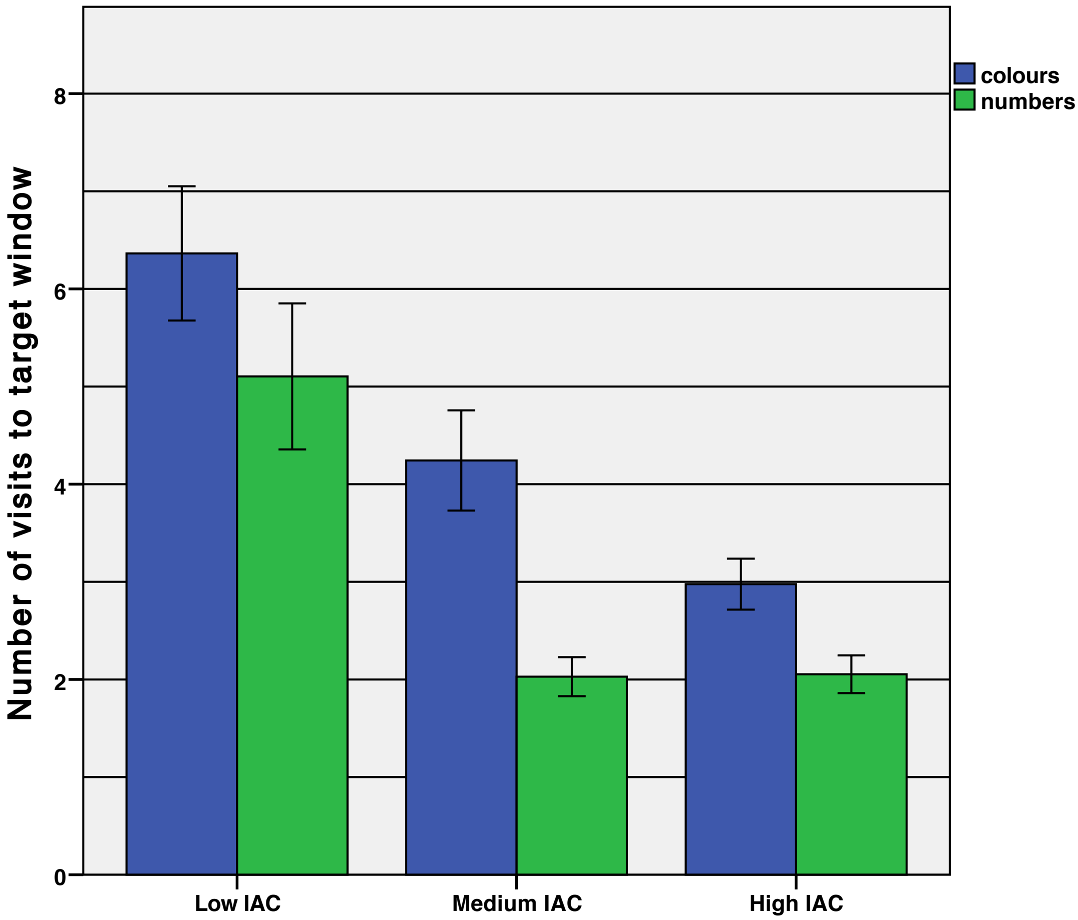


Figure 4.2: The interaction between block type and IAC for number of visits to the target window. The error bars represent ±1 standard error.

Duration of first visit to target window

There was no significant main effect of block type on the duration of the first visit, $F(1,30) = 3.05$, $p=.09$, $\eta^2 = 0.09$. Participants looked longer at the target source as IAC increased from Low to High. Post-hoc comparisons showed that participants looked longer in the High-IAC condition (M=1.84, SD = 1.35) than in the Low/Medium-IAC conditions, ps <.001. However, there was no difference in duration between the Low-IAC (M = 0.45, SD = 0.46) and the Medium-IAC (M = 0.05, SD = 0.04) conditions, p=.47. There was a significant interaction effect between IAC and block type, $F(2,30) = 5.70$, $p<.01$, $\eta^2 = 0.28$ (see Figure 4.3). There were no difference between block types in the Low-IAC condition, $t(10) = -1.86$, $p = 0.09$, nor the Medium-IAC condition, $t(9) = -0.29$, $p = 0.7$. However, in the High-IAC condition, participants looked significantly longer for colours (M = 2.18, SD = 1.59) than numbers (M = 1.49, SD = 1.01), $t(11) = 2.76$, $p = 0.02$.

Blocks placed after first visit

People placed more blocks correctly after the first visit for numbers (M = 4.77, SD = 2.33) than colours (M = 3.08, SD = 1.81), $F(1,30) = 63.86$, $p<.001$, $\eta^2 = 0.68$. They also placed more blocks as IAC increased, $F(2,30) = 12.54$, $p<0.001$, $\eta^2 = 0.46$. Tukey post-hoc comparisons show there was a difference between the Low IAC and Medium/High IAC conditions (ps<.01), but not between Medium and High IAC conditions (p=.77). There was a significant interaction effect between

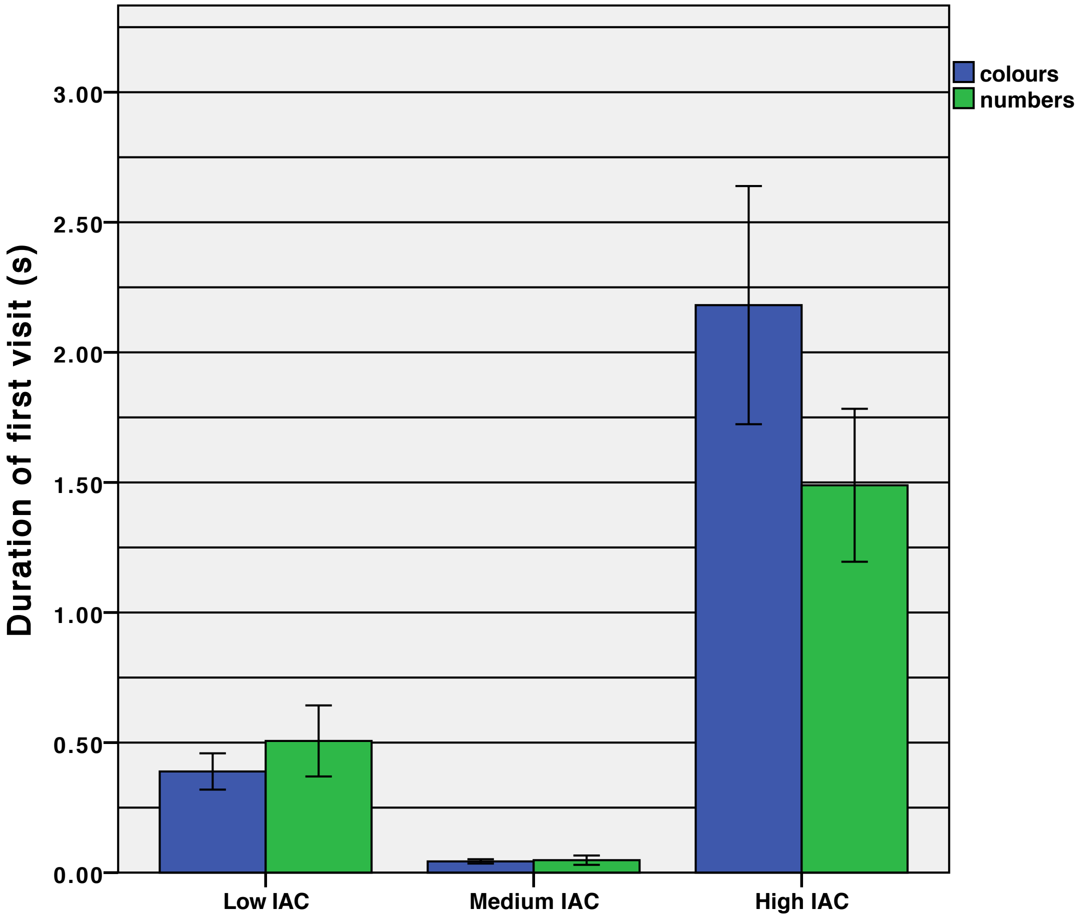


Figure 4.3: The effect of IAC on the duration of the first visit to the target window. The error bars represent ±1 standard error.

IAC and block type, F(2,30) = 8.96, p<.01, η^2 = 0.37 (see Figure 4.4). When IAC was Low, the number of blocks that were copied correctly after the first visit did not differ significantly for colours or numbers.

Global task performance

The interactions between block type and IAC on global task performance measures all had the same trend: people performed the same for colours and numbers when IAC was Low, but differences appeared between the block types as IAC increased. As this trend was the same for each interaction, the statistical results of the interactions are reported but their specific trend will not be repeated.

Number of incorrectly placed blocks

Participants placed more blocks incorrectly for colours (M = 0.54, SD = 0.45) than numbers (M = 0.32, SD = 0.21), F(1,30) = 10.71, p=.003, η^2 = 0.26. As IAC increased and participants were keeping more items in memory, they increasingly placed more incorrect blocks, F(2,30) = 14.71, p<.001, η^2 = 0.50. Tukey post-hoc comparisons show there was a difference between the Low IAC condition (M = 0.16, SD = 0.18) and Medium/High IAC conditions (ps<.01), but

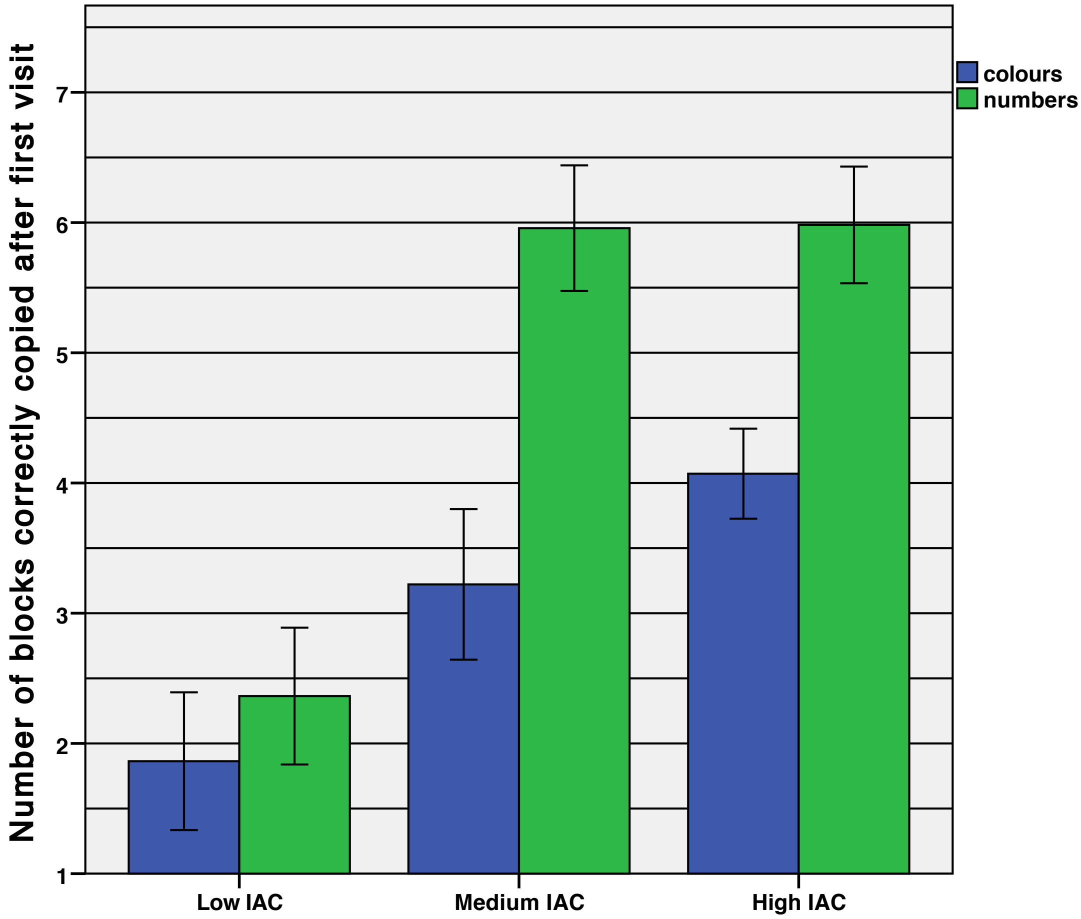


Figure 4.4: The interaction between block type and IAC for number of blocks correctly placed after the first visit to the target window. The error bars represent ±1 standard error.

not between the Medium (M = 0.49, SD = 0.35) and High IAC conditions (M = 0.63, SD = 0.36) (p = .3). There was a significant interaction effect between IAC and block type, F(2,30) = 3.36, p<.05, η^2 = 0.18. When IAC was Low, the number of blocks that were copied incorrectly did not differ significantly for colours or numbers, but as IAC increased, participants placed more blocks incorrectly for colours.

Number of incorrectly submitted trials

The number of trials that were submitted incorrectly was generally low, but participants submitted more incorrect trials for colours (M = 0.1, SD = 0.16) than numbers (M = 0.04, SD = 0.08), F(1,30) = 5.28, p=.029, η^2 = 0.15. There was no significant effect of IAC, F(2,30) = 2.70, p=0.08, η^2 = 0.15, nor any interaction, F(2,30) = 1.65, p=.2, η^2 = 0.10.

Trial time

Two trial completion times are considered here: total time and time excluding lockout. Looking at the actual completion time, participants took longer to complete a trial when they were copying colours (M = 25.80, SD = 7.06) compared to when copying numbers (M = 22.24, SD = 4.47), F(1,30) = 44.08, p<.001, η^2 = 0.60. As IAC increased from Low to Medium to High, participants took longer to complete a trial, IAC, $F(2,30) = 15.91$, $p<0.001$, $\eta^2 = 0.52$. Tukey post-hoc comparisons show there was a difference between Low/Medium and High (ps<.01), but not between Low and Medium ($p = .12$). There was a significant interaction effect between IAC and block type, $F(2,30) = 11.05$, $p<.001$, $\eta^2 = 0.42$.

With the lockout time in the High-IAC condition removed, the same effects were found for block type, $F(1,30) = 34.55$, $p<.001$, $\eta^2 = 0.54$, and IAC, $F(2,30) = 8.18$, $p=0.001$, $\eta^2 = 0.35$. Tukey post-hoc comparisons show there was still a difference between Low and High ($p=.001$), but no longer between the Medium IAC and Low IAC or High IAC conditions (ps >.1). There was a significant interaction effect between IAC and block type, $F(2,30) = 8.13$, $p=.002$, $\eta^2 = 0.35$.

Qualitative data

The screen recordings from the eye-tracker were played back to further investigate people's behaviour. Although this helped understand some behaviour which could not be determined from the quantitative data alone, these observations only serve to explain some of the quantitative measures and are not the main focus of the analysis.

The visit durations in the Medium IAC condition were suspiciously short. Upon replaying the screen recordings, it appeared that participants often accidentally moved their cursor over the grey mask of the target source. This was counted as a visit by the program, even though participants may have not intentionally moved their cursor to this part of the screen to look at the target source. They did not spend a long time looking at the target window, but also did not immediately move blocks either, and sometimes waited multiple seconds before they made a move.

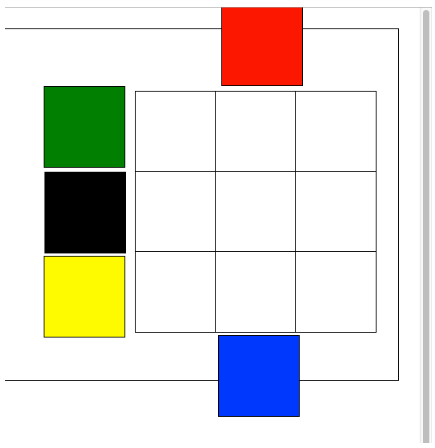


Figure 4.5: Participants placed blocks outside of the output window as 'placeholders'.

During the 1-s lockout in the High IAC condition, participants changed their minds about visiting the target window on numerous occasions. They placed their mouse cursor on the mask, but left this field before it was uncovered to move one or more blocks. It could be this decision also occurred in the Medium IAC condition, but as there was no lockout the mask was already uncovered before people made this decision, and would explain the very short visits.

People sometimes placed the blocks as 'placeholders' as shown in Figure 4.5: they placed several blocks outside of the output window next to the position they thought it belonged to, but did not place it there yet. Only after viewing the target again, they placed the blocks in the output window. Looking at quantitative data alone, this type of strategy would be depicted as one long view at the target, after which all blocks were placed in one go. This is true to some extent, but as people could already place the blocks and offload their memory without this being recorded by the program, they only had to check if this position was correct on the subsequent visit, and is different from a strategy where people spent a long time trying to memorise the blocks after which all blocks were placed.

4.1.4 Discussion

This study replicated the Blocks World Task with the aim to see if the effect on IAC on a copying task, as previously found when using colours, can be extended to copying numbers, and if a deeper encoding of numbers into memory makes people more accurate.

The main findings are:

- Numbers are easier to copy than colours, but only if IAC is increased
- Increases in IAC make people adopt memory-intensive strategies
- Increases in IAC increases errors

The effect of block type on task strategies and performance

Changing the block type from colours to numbers made it easier to memorise the blocks: people made shorter and fewer visits than when copying colours, but were still able to place more blocks. This difference in performance is only found in the Medium and High IAC conditions, which further suggests the difference between numbers and colours is due to the memorability of the information, and the interactions on most of the dependent variables further show that the difference between block types was dependent on the level of IAC.

The difference between numbers and colours fit well with the distinction in Baddeley's (**?**) seminal model of working memory between processing visuo-spatial and verbal information. As numbers can be rehearsed and therefore refreshed in working memory, they are likely easier to memorise. After the study had ended, several participants in this study explained they used some sort of rehearsal and tried to say the numbers out loud in their head. For colours, some indicated they tried to either capture a mental picture of the colours in their mind, while others rehearsed the word for each colour (e.g. 'red' and 'blue') but they felt they were able to remember fewer items using this strategy. This is consistent with Chincotta, Underwood, Ghani, Papadopoulou, & Wresinski (1999)'s finding that people have a greater memory span for Arabic numerals (e.g. 1, 2, 3) than corresponding words (e.g. one, two, three).

Though an increase in IAC made people rely on memory and people seemed to memorise numbers more easily than colours, an increase in IAC did not make people more accurate in copying numbers as in Gray & Fu (2004), where people had to memorise VCR programming information. The error rate for numbers was however already quite low in all conditions, and moreover people were not explicitly instructed or trained to memorise the information before each trial as in Gray & Fu (2004).

The effect of IAC on task strategies and performance

The effect of IAC on people's cognitive strategies, as found in previous studies, is replicated in this study: as the cost of accessing the target window increased, people increasingly relied on memory (Gray et al., 2006; Morgan et al., 2009; Waldron et al., 2007). People switched from a perceptual to a memory-based strategy by making fewer but longer visits to the target window and placing more blocks immediately after the first visit. This strategy worsened their global performance as they took longer to complete the task and placed more incorrect blocks throughout the trials.

An increase in IAC also made people rely more on memory when copying numbers. When a mask was added over the target window, people visited the target window less often and they placed more blocks after the first visit. In both the Medium IAC and High IAC conditions, people on average placed around six blocks correctly after the first visit, after which they needed one more visit to look at the last three blocks. The average number of items people can memorise is 7±2 items (**?**), and potentially six blocks was the maximum that people could reasonably memorise.

In previous studies, adopting a memory-based strategy when copying text and numbers improved accuracy (Gray & Fu, 2004; **?**). In the current study, the hypothesis that memorising numbers due to an increased IAC would make people more accurate is not supported. The error rate was overall low and upon reflection the interaction of moving blocks may have made people sufficiently slow to hardly make any errors. In Gray & Fu (2004) and **?**, people typed in data using a computer keyboard.

With the BWT, it was difficult to measure visits to the target window in the same manner for all conditions. For the Low IAC conditions, eye fixations were used, whereas for the Medium and High IAC conditions, uncoverings of the mask were used. This introduced several problems. First, while eye-tracking measures show how long and how often people are looking at a particular part of the screen, it can not reveal if people are actually perceiving or processing the data that is displayed (Waldron et al., 2007). For the Medium and High IAC conditions, an interaction was required and a conscious decision had to be made to reveal the target in these conditions. It would therefore seem likely that uncoverings more reliably measure visits to the target window. However, the uncoverings for the Medium IAC conditions were suspiciously short. Playing back the screen recordings suggests participants often accidentally uncovered the window, and it is therefore less clear if these are a reliable measure of actual visits.

Summary

The purpose of this study was to study the effect of IAC on strategy, speed and accuracy when copying from one source, and investigate if the effect of IAC, as found in previous experimental studies, would extend to copying numbers. In order to be able to compare results of this study with previous IAC studies the same task paradigm of the Blocks World Task was used.

The hypothesis that people would need fewer switches and would perform better for numbers than colours is supported, indicating that numbers are easier to memorise. The hypothesis that memorising numbers due to an increased IAC would make people perform better is not supported. The error rate was overall low and upon reflection the interaction of moving blocks may have made people sufficiently slow to hardly make any errors. Furthermore, in Study 1 people saved up data to enter a lot in one sequence, and errors increase as people have to copy more (Healy et al., 2004). Potentially the setup of this experiment was not suitable to reliably measure an increase in errors.

The information was copied from one source, and for each participant the IAC was the same throughout the experiment: it was either low, medium, or high. Therefore, while using the BWT task paradigm was appropriate in comparing its results with previous IAC studies, it did not resemble the data entry tasks observed in the financial office setting of Study 1.

Considering these limitations, it was decided after this study to not continue with the same task paradigm, but certain learning points can be taken away from this study.

First, the entry method matters. An entry method that is sufficiently slow may make people accurate, regardless of the level of IAC. Second, the effect of IAC generalises, but it is not clear how the results of this experimental paradigm can be translated in suggestions in how people dealing with differing levels of IAC can best be supported in their work. In contrast to this study, where information came from one source, and always had the same level of IAC for each participant, Study 1 showed that for entering expenses, people have to enter different types of information, these do not all come from the same source, and each source can have a different access cost.

Third, the measured short visits in the Medium IAC conditions may have confounded the results. In future studies, the setup of the experiment should be designed so that participants do not accidentally access the source when they do not intend to.

Lastly, for each dependent variable, the same consistent measure should be used to compare across conditions. This could be data provided by an eyetracker, or interaction measures such as mouse clicks, key presses or interkey intervals.

The next study will try to more closely simulate the expenses task observed in the financial office setting of Study 1 and 2. People will have to enter and collect information from multiple sources with different IACs. The aim is to see how these differences in IAC influence people's switching strategies between entering and looking up data when copying from multiple sources.

4.2 Study 4: Copying data from multiple sources

4.2.1 Introduction

Study 3 showed that as it takes longer to access an information source needed for a copying task, people spend a longer time looking at that source. They try to group and memorise as much information, in order to minimise the number of revisits to this source. Though people copied over more items after one visit using this strategy, they also made more errors overall. In the experiment, all information was to be found on a single source. People may have tried to memorise too much items in one visit, and upon entry relied on incorrectly memorised information in the head.

Data entry in office workplaces often does not involve a single source, but information can be spread over various sources. These sources are often not all equally easy or hard to access. How do people prioritise how they enter and look up information from these different sources? Do they enter the easy items first? And will they still try to group and memorise as many items? Or will they look up and enter items from one source at a time? In order to support people in looking up information for data entry work, it is important to first understand how people currently manage these information tasks.

In the contextual inquiry of Study 2, participants started data entry by first collecting all physical sources first and placing these on their desk. They then entered this information, which was nearby, first. They postponed accessing other sources until they needed to enter it during the task. If information took too long to access, they would set aside the task and continue with other tasks. One of the factors that influenced the different strategies appeared to be the time cost to retrieve the information.

This study tests the assumption that observed strategies from Study 2 are influenced by the time costs to access information sources. Whilst prior studies have demonstrated that various tasks can involve the use of multiple information sources (Murphy, Chen, & Cossutta, 2016; Su, Brdiczka, & Begole, 2013; **?**), there are limited studies that measure how people use these sources, and to what extent the time cost to access a source influences these decisions.

The following questions will be addressed:

The soft constraints hypothesis predicts that people choose and adapt their task strategies in order to minimise time (Gray et al., 2006). Study 3 found that the longer it takes to access information, the more items people try to memorise in one visit. Based on these findings, the following hypothesis is made:

H1. As IAC increases, people will try to group and memorise multiple items to minimise visits.

In O'Hara & Payne (1998)'s study on the effect of IAC on problem solving tasks, participants in Low-IAC and High-IAC conditions initially performed the same type of strategies. However, over time participants in a High-IAC condition learnt more efficient strategies, whilst participants in a Low-IAC condition continued to use the same strategy. Prior work has also shown that people who are exposed to High-IAC situations will continue to use the strategy they learnt to be the most efficient, even in situations where the cost to access information is no longer high (Patrick, Morgan, Tiley, Smy, & Seeby, 2014). It is therefore expected that once participants learn it is more efficient to group High-IAC items, they may adopt this strategy for Low-IAC items as well. In Study 2 of this thesis, people tried to enter items that were nearby in the environment first, and postponed looking up other information until later. Based on these findings, the following hypothesis is made:

H2. As the experiment progresses and people become aware how costly it is to access certain sources, they will learn and choose to enter all the Low-IAC items first, in a batch, and then the High-IAC items second, also in a batch, rather than looking up each item as they need it.

Lastly, if people group items and spend a longer time looking up information, this means they will be interrupted from their data entry task for a longer time. The longer people are interrupted from a primary task, the slower they are to resume that task after the interruption (Monk, Trafton, & Boehm-Davis, 2008), and the more likely they are to make resumption errors (e.g. Brumby et al., 2013). Based on this, the following hypothesis is made:

H3. As IAC increases, people will be slower and make more errors.

4.2.2 Method

Participants

Thirty-three participants (12 male) ranging from 18-52 years (M = 26, SD= 8) took part in the experiment. They were recruited from a university subject pool and received £4 for their participation.

Task

The experimental data entry task was framed around the expenses task from Studies 1 and 2. For this task, the user has to complete a number of data entries regarding incurred expenses in order to get the expenses reimbursed. They enter this into a claim form, which looks similar to a spreadsheet. Users can choose to either fill in multiple expenses in one sheet, in which each row corresponds to one expense, or have separate spreadsheets for each expense. The current experiment will use the scenario where users enter multiple expenses in one sheet.

For each trial, participants were presented with a data entry sheet consisting of two expense claims (see Figure 4.6). They had to complete each row by entering a financial amount to specify an expense that was made, and an account code to specify which account the expense belonged to. The retrieved these data items by switching to two other pages. One page contained the amounts, and another page contained the account codes. The participant could go to a page by clicking on the corresponding name in the horizontal menu at the top of the screen. Only one page could be viewed at a time and covered the full screen.

Materials

The numbers to be entered were made to resemble values that are ecologically relevant to the task. The account codes were similar to codes that are currently used by one of the universities studied in Chapter 3, and have a fixed length of six digits (e.g. 654273). Office workers at this university usually worked with the same 20 or 30 account codes. They were aware they worked with the same set of codes, but still had to look a code up each time they needed to enter it, as the codes were too difficult to enter from memory. The codes had a length of six digits, and the

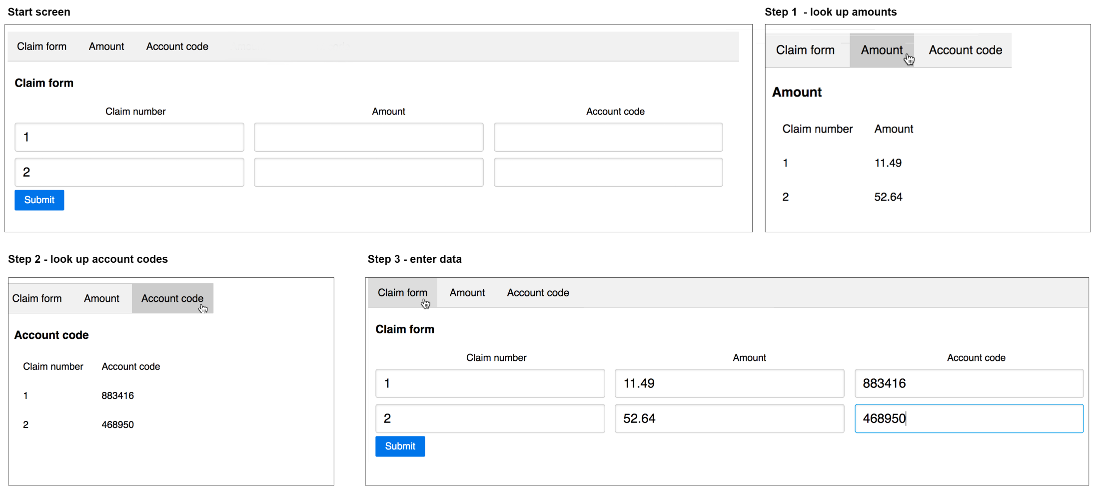


Figure 4.6: The data entry task. At the start of each trial, participants were presented with a data entry form with two expense claims, and had to enter four data items in a data entry form. The data items were retrieved from a separate Amounts page (Step 1) and an Accounts page (Step 2) and entered into the data entry form (Step 3).

string was random with no pattern. Amounts consisted of two digits on the integer part and two digits on the fraction part (e.g. 11.95).

The experiment was conducted in a maximised web browser on a desktop computer with a 24-inch monitor and a resolution of 2048x1152 pixels. Participants used a computer mouse and number keypad, and the option to copy and paste information was disabled. If the participant switched from the data entry form to another page and back, the cursor stayed in the same data entry field. The task interface was developed in HTML, CSS, JavaScript and PHP. All mouse clicks, key presses and timestamps were recorded using JavaScript.

Design

A between-participants design was used with one independent variable, the presence or absence of a delay when switching to one of the information pages. In the Control condition, there were no delays in switching between any of the pages. In the High-Amount condition, there was a 2-s delay when switching to the Amount page, and in the High-Account condition there was a 2-s delay when switching to the Account page. There were no delays in switching back to the data entry form. On a trial-by-trial basis, the main dependent variable was whether people interleaved or not: did participants enter the data items in sequential order, or did they interleave between the two expenses? Two values had to be entered for each expense: an amount and an account code. If participants entered the amount and account code of one expense before entering the other expense, this was considered a sequential order. If participants entered amounts of each expense first, followed by entering the account codes or vice versa, this was considered interleaving. All key presses were recorded to determine in which order data was entered. Page switches were recorded to capture the number and duration of switches to information pages. Other dependent variables were trial completion time and data entry error rate. In addition, we analysed the type of errors made to determine whether participants made more omission errors in any of the conditions.

Procedure

The experiment took place in a closed quiet room. It was explained to participants that the task involved entering expenses, and that for each trial they had to enter two expenses. They were not advised to use a particular strategy, but it was explained it was important to complete all data entry fields before proceeding to the next trial, as they could not return as soon as they had pressed 'Submit'. There were no restrictions in the number or duration of times they could switch between pages, or the order in which they completed the trial. One trial consisted of two expenses, i.e. four data entries. Participants first completed two practice trials to familiarise themselves with the task, and were free to ask any questions; data from these trials were not included in the analysis. After that, the experimental session consisted of 50 trials, divided into 5 blocks of 10 trials. After each block, there was an opportunity for the participant to take a short break. A prompt appeared on the computer screen, and the recording time was paused. Participants could carry on with the experiment by pressing a button on the screen. For each block, a set of 20 different amounts and 20 different account codes were used. These sets were re-used for every block, so in total, each number was presented five times throughout a session. The experiment took approximately 30 minutes.

Pilot study

Two pilot studies were conducted with colleagues of the researcher to test the experimental design. In particular, the pilot studies aimed to see if the length of the experiment was long enough for participants to learn and develop strategies, but not too long to tire the participant.

During the pilot studies, there was a scheduled break after every 5 trials. Both participants mentioned the break prompts happened too frequently, and experienced them as disruptive. They did not find the experiment too long. One participant could not remember which computer tabs had an increased IAC. As a result, he did not adapt his strategies according to anticipated IACs and entered the data items row by row. The second participant mentioned that the increased IACs definitely made her more careful in checking the numbers were correct. The participants were aware some of the numbers occurred more than once, but the numbers did not occur often enough to be able to memorise them.

For the real experiments, the breaks were reduced to happen after every 10 trials. In addition, the names of information pages with an increased IAC were underlined in the horizontal menu. This visual feature was added to help users see more easily which windows had a delay.

Data analysis of data entry strategies

A bottom-up approach was taken to group and analyse people's data entry strategies. For the first iteration of grouping, each trial was grouped into one of two categories: a sequential or interleaving category. If participants first entered the amount and account code of one expense before entering the other expense, this trial was grouped in the sequential category. If participants entered amounts of each expense first, and then entered account codes, or the other way around, this trial was grouped in the interleaving category. On a small subset of trials (<1%) neither of these strategies was chosen: for example, participants first entered the amount of one expense, followed by the account code of the second expense. These trials were also grouped in the interleaving category, as participants switched to entering the second expense before completing the first expense.

Mouse clicks to switch between pages were used to code the order of people's actions, with the aim to capture people's switching behaviour between entering and looking up information. During the second iteration of grouping, for each trial the order of actions was considered and

Condition	Interleaving rate	Number of visits	Duration of visits (s)	Error rate	Trial completion time (s)	Trial completion time (excl lockouts) (s)
Control	31.17% (42.24%)	5.06 (1.03)	1.83 (0.80)	8.68% (10.91%)	29.99 (10.94)	29.99 (10.94)
High-Amounts	34.18% (41.5%)	5.24 (1.81)	1.60 (0.34)	3.77% (2.80%)	28.88 (6.92)	24.42 (5.91)
High-Account	73.2% (41.1%)	4.29 (0.45)	1.78 (0.60)	5.18% (4.13%)	29.23 (7.44)	25.05 (7.32)

Table 4.2: The means (and standard deviations) of all dependent measures for each condition. The rates are calculated by dividing the number of occurrences to the number of opportunities, e.g. an interleaving rate of 50 percent means participants interleaved on 50 percent of trials.

the trial was either grouped under a new strategy group for this order, or the trial was grouped under an existing strategy group.

4.2.3 Results

Table 4.2 summarises the results of the dependent measures for the three conditions. The distribution of dependent measures were skewed, so non-parametric Kruskal-Wallis tests were used to analyse effects of IAC on the dependent variables. A p-value of 0.05 was used for assessing the significance of all statistical tests.

Interleaving strategies

A trial was labelled as 'interleaving' if the participant started entering one expense but interleaved to the other expense before completing the first one. The interleaving rate for each condition was calculated by dividing the number of trials where people interleaved by the number of total trials.

The boxplots in Figure 4.7 show the variability of interleaving rates across conditions. The Control condition had a median interleaving rate of 6%, the High-Amount conditions had a median interleaving rate of 12%, and the High-Account condition had a median interleaving rate of 96%.

Participants interleaved most often between expenses in the High-Account condition (M = 73.2%, SD = 41.1%), compared to the Control (M = 31.17%, SD = 42.24%) and High Amount (M = 34.18%, SD = 41.5%) conditions, $\chi^2(2) = 6.81$, p = 0.03. A post-hoc Dunn's test showed there was a difference between the High-Account condition and the Control (p = 0.02) and the High-Amount p = 0.03) conditions, but not between the Control and High-Amount conditions (p=0.9).

In the High-Account condition, participants predominantly switched to the page with the Amounts first, which had no time delay, and entered these into the data entry form. In the other two conditions, participants mostly entered an amount and account code of the first expense first, and then entered the amount and account code of the second row.

Across conditions, most participants were consistent in their strategy choice, and either interleaved between expenses on almost no (0%) or all (100%) trials. This is illustrated in Figure 4.8, which shows the distribution of interleaving rates for each condition. The lines all have peaks

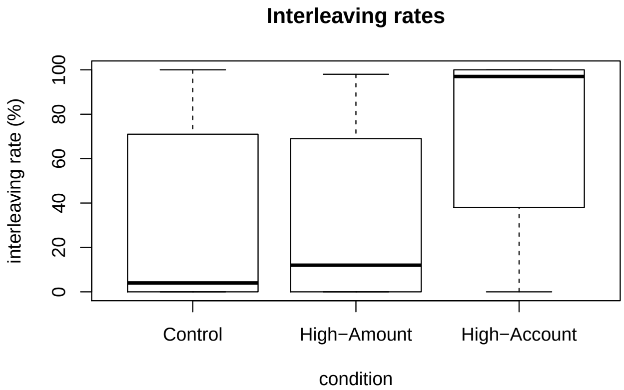


Figure 4.7: Boxplot of interleaving rates in each condition.

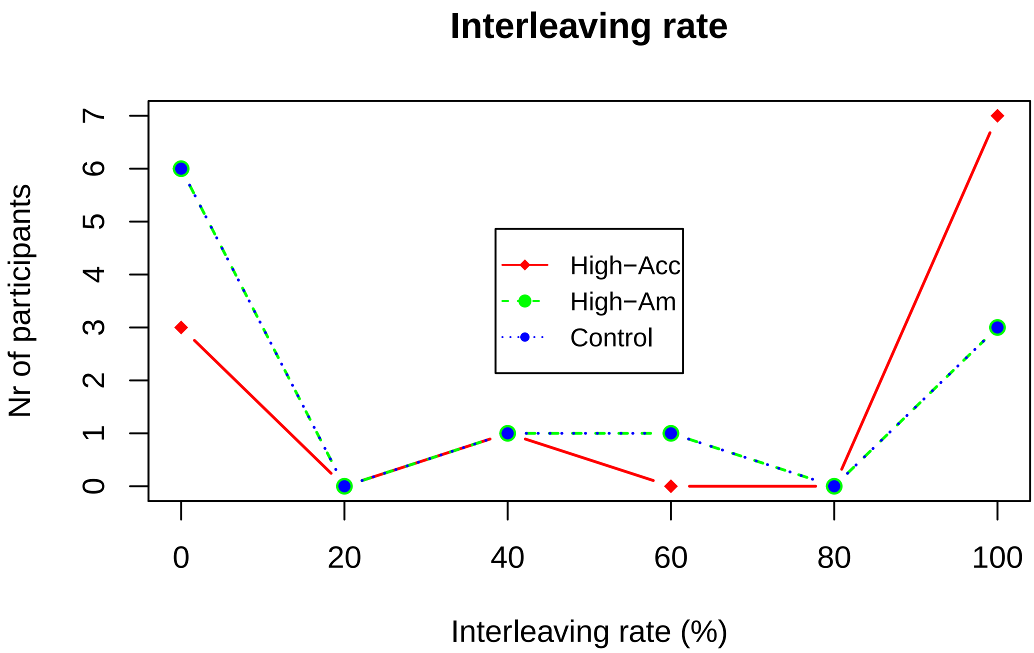


Figure 4.8: Line graph showing the frequency of interleaving rates for each condition. As can be seen, all three lines have two peaks at 0 and 1, which means that most participants interleaved on 0% or 100% of all trials.

at the left and right end, indicating the interleaving rate was predominantly 0 or 100% in each condition.

The consistency in strategy choice is further illustrated in Figure 4.9, which displays a plot for each participant across trials. The x axis plots the trial number, and the y axis displays whether they interleaved on that trial or not: a value of 0 means they did not interleave, and a value of 1 means they did interleave. These plots further illustrate that most participants were consistent in interleaving on no or all trials (see for example Participant 6, who interleaved on all trials). A subset of participants switched between strategies at the first couple of trials before sticking with one strategy, such as Participants 9 and 12. Lastly, participants 29, 32 and 33 seemed to switch between the strategies throughout the experiment and did not stick with a particular strategy.

Number and duration of visits

There was no difference in the number of visits, $\chi^2(2) = 2.90$, $p = 0.23$. On average, participants made 4 visits per trial (i.e. one visit per data entry). Participants visited an information page for 1.8 seconds on average, and there was no significant difference in duration of visits between conditions, $\chi^2(2) = 0.30$ $p = 0.8$.

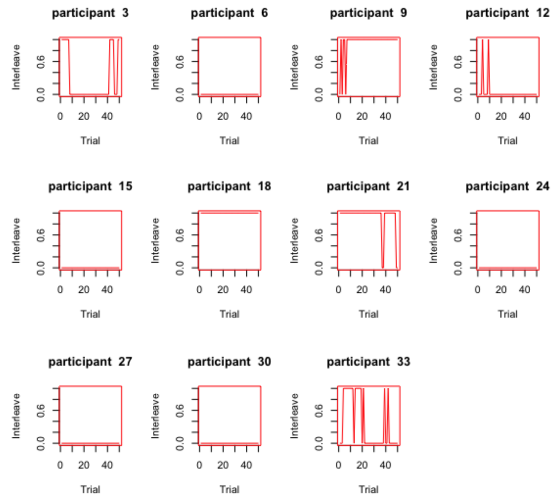
To get a better insight in the specific order in which participants viewed and entered items, the trials were grouped based on the order of actions. There were eight different possible actions: viewing the first amount (V-Am1), viewing the second amount (V-Am2), viewing the first account code (V-Acc1), viewing the second account code (V-Acc2), entering the first amount (E-Am1), entering the second amount (E-Am2), entering the first account code (E-Acc1), and entering the second account code (E-Acc2). This iteration of grouping the trials resulted in 17 different strategy groups in total, with the majority of trials (98%) grouped in the same four groups, which are shown in Figure 4.10. For example, the first strategy (a) shows a strategy where participants started a trial by visiting the Amount page, and then visiting the Accounts page. They then entered both the amounts of the first expense (Am1) and the account code of the first expense (Acc1). They then visited the Amounts page again, and entered the amount of the second expense (Am2), and then visited the Accounts page again and entered the account code of the second expense (Acc2). Table 4.3 shows the frequency with which these strategies were chosen per condition.

	Row		Column			
Condition	First row (a)	First & Second row (b)	Amounts (c)	Amounts & Accounts (d)	Other	Total
High-Account	34% (48)	4% (6)	57% (80)	2% (3)	3% (4)	100 (141)
High-Amount	20.99% (44)	16.57% (35)	49.72% (104)	9.39% (20)	3.31% (7)	100 (210)
Control	11.2% (16)	21.6% (32)	54.4 (81)	12% (18)	0.8% (1)	100 (148)
Total	21.18% (190)	15.02% (73)	52.96% (265)	8.37% (41)	2.46% (12)	100 (499)

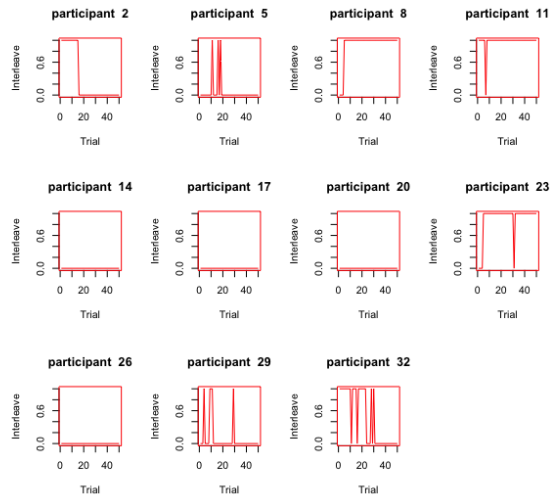
Table 4.3: The most common grouping strategy was to chunk the items into three groups. The strategies are shown graphically in Figure 4.10.

Errors and trial completion time

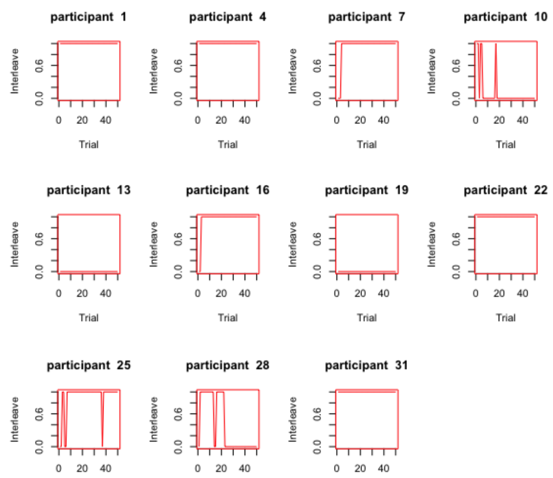
There were 200 data entries, so in total there were 200 opportunities for a participant to make a data entry error. The error rates were calculated as the number of errors divided by the number of entries. Though the mean error rate was higher in the Control condition (M=8.68%,



(a) Participants in the Control condition.



(b) Participants in the High-Amount condition.



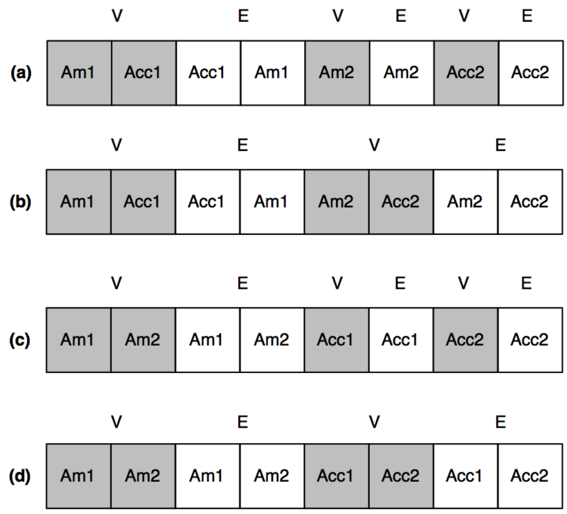


Figure 4.10: The sequence of the most common grouping strategies. V = visit to the data source, E = entry of the data item. For example, in Strategy (a) a participant first viewed Amount1 and Account1, and then entered Amount1 and Account1. He/she then viewed Amount2 and entered it, and then viewed Account2 and entered it.

SD=10.90%) compared to the High-Amount (M=3.77%, SD=2.79%) and High-Account (M=5.18%, SD=4.13%) conditions, this difference was not statistically significant, $\chi^2(2) = 0.41$, p = 0.8. The High-IAC conditions already had an extra time cost to overall completion time, due to the delay to one of the pages. Therefore, two completion times were calculated: one of the actual completion time, which included the lockout times, and another with these times removed. Considering these two times, there was no difference in the time it took to complete a trial using the actual completion time, X(2) = 0.15, p= 0.9, and with the lockout times removed, $\chi^2(2) = 2.92$, p = 0.2. On average, participants took about 29 seconds per trial across conditions.

? taxonomy of number entry errors was used to analyse the types of data entry errors that were made. As can be seen in Figure x , the most prominent error types were when participants had a digit(s) wrong (60 times), when a data entry was skipped (75 times) or when they entered a correct number, but in the wrong input field (57 times): these types of errors make up for 61% of all errors.

Qualitative findings

After the experiment had ended, participants were debriefed and the purpose of the study was explained. Some participants recalled their strategies and gave additional explanations behind them.

x participants adapted their strategy several times throughout the experiment, in order to find the quickest way to complete the task. Because amounts were shorter and easier to remember, several participants mentioned they tried to first view all amounts before entering them. They tried this strategy with account codes as well, but these were longer and therefore it was more difficult to memorise two at a time. As. a result, most participants ended up viewing and entering each account code one by one.

Several participants noticed that numbers re-occurred throughout the experiment. They felt it was easier to memorise a number that had already occurred earlier in the experiment, so when a trial contained a number they had entered before, they would use the 'Grouped' strategy: they viewed another window and item before returning to the entry form. Furthermore, some participants started a trial by not viewing any other windows, but re-entering the data items of the previous trial in the data entry form. They would then visit the information source windows to check if the items were the same, or if they needed to change some of the items.

4.2.4 Discussion

The aim of this study was to investigate the effect of IAC on how people manage task interruptions to look up information from multiple information sources with varying IACs. IAC influenced the order in which people entered data, and whether and how they chunked data items. In the Control condition, when there were no differences in IAC, participants tended to complete a data entry sheet in sequential order, and completed one expense before moving to the next one. When comparing this with the High-IAC conditions, people interleaved significantly more between expenses in the High-Account but not High-Amount condition. These findings partly support the hypothesis that people postpone looking up information with a higher IAC, but it does not account for why people continued to enter the sheet in sequential order in the High-Amount condition.

These results can be explained when considering the order in which the data was presented, and the order in which items were entered. Across conditions, participants predominantly started each trial by entering the first cell of the data entry sheet, the amount of the first expense, regardless of whether the Amounts page had a 2-s delay. However, the second item they entered was dependent upon which window had a delay: if the amounts window had a delay, they would enter the account code first. If there was a delay with the accounts, they would enter the second amount first. It seems that IAC does not influence the first visit, but does affect subsequent visits. Even though the IAC was consistent throughout the experiment, potentially the experiment was too short for participants to learn which of the windows had a delay and only adapted their strategy after they had already entered the first item. Furthermore, participants tended to stick to the same strategy they had started with throughout the experiment. We will discuss this result further in the General Discussion.

In Experiment 1, both expenses were shown on the same page, and could be seen as part of the same task. We wonder whether differences in IAC also makes people interleave between two distinct tasks, which are shown on separate pages. The aim of the next experiment was therefore to study the effect of differences in IAC in a multi-task setup.

Ordering strategies

IAC had an effect on people's ordering strategies, but only in the condition where the IAC was increased for accessing account codes. People first looked up and entered the low IAC items, the amounts, before they entered the high IAC items, the account codes. However, if the cost to access the amounts was increased, people entered items row by row, as participants did in the control condition.

In O'Hara & Payne (1998)'s study on the effect of IAC on problem solving tasks, participants in Low-IAC and High-IAC conditions initially performed the same type of strategies. However, over time participants in a High-IAC condition learnt more efficient strategies, whilst participants in a Low-IAC condition continued to use the same strategy. Prior work has also shown that people who are exposed to High-IAC situations will adopt and continue to use a certain strategy, even in situations where the cost to access information is no longer high (Patrick et al., 2014). It is therefore expected that once participants learn it is more efficient to group high-IAC items, they may adopt this strategy for low-IAC items as well. In Study 2 of this thesis, people tried to enter items that were nearby in the environment first, and postponed looking up other information until later. Based on these findings, the following hypothesis is made:

In previous studies, people would postpone or not address information needs if there was a high time cost associated with it (Sohn et al., 2008). This was also found in Study 2. People initially tried to do tasks as they needed it, but if it took too long, they would put it aside.

In the current study, this was also found but the cost to memorise an item overrode the cost to access it: people would first enter the amounts, even if the cost to access them was higher. If the cost was high, they would try and memorise them both before entering them. This shows people try to minimise interleaving between viewing and entering information.

Grouping strategies

The soft constraints hypothesis predicts that people choose and adapt their task strategies in order to minimise time (Gray et al., 2006). Study 3 found that the longer it takes to access information, the more items people try to memorise in one visit. Based on these findings, the following hypothesis is made:

Performance

Grouping strategies

IAC also had an effect on how often people switched between visiting and entering information. In the conditions where the amounts were low, people often visited multiple items before returning to the entry interface. Though there was no difference in errors in the current study, it can be imagined this strategy is more risky as people are away for longer so resuming a task is more difficult, and they are storing more information in memory, so people may misremember information.

In the condition where the IAC for amounts was high, this strategy was not used as frequently. IAC thus does not have to be bad if the information is easy enough to briefly hold in memory (condition amounts high).

IAC made people changes strategies even if only two of the four information sources had an increased IAC. This is in accordance with (**?**), who showed that a more memory-based strategy can be trained for future situations. When people were exposed to an interface with an increased access cost, they adopted a memory-based strategy and retained this strategy, even when they then interacted with an interface with lower access costs.

Conclusion

This study investigated the effect of IAC on looking up information from multiple sources, and showed that, if all information has the same IAC it is better if this is low, differences in IAC between sources can make people schedule their subtasks more efficiently and effectively. People needed less visits and made fewer errors. However, for this study people only had one data entry task at a time to focus on, and the sources only contained information for this single task. In Study 1 and 2, people often batched and saved up multiple data entry tasks, and sources included more than just the information for one task: for example, if people had to look up an account code, this was often retrieved from a spreadsheet with all account codes. They thus not only had to manage subtasks of one data entry task, but had to coordinate multiple data entry tasks. In order to save visits, people may therefore look up information for several tasks and interleave. One of the participants from Study 1 shared that this did occur and said she and colleagues had to be very careful to not input information in the wrong form. If people group information subtasks per IAC for a single task, would they also group these subtasks per IAC when dealing with multiple tasks? In order to answer this question, a follow-up study was run to see the effect of IAC on interleaving behaviour between two claim forms.

4.3 Study 5: Interleaving between data entry tasks

4.3.1 Introduction

Study 4 showed that people group subtasks of looking up information according to the information's IAC: if all information sources were equally easy to access, participants looked up and entered information for a task in sequential order. However, if some sources were more costly to retrieve, people first retrieved and entered all low IAC items, before they retrieved high IAC items. If people deal with multiple data entry tasks, would this also make people more likely to look up and enter low IAC items across tasks, rather than completing one task at a time? In Study 2, participants saved up data entry tasks to do in one session and occasionally had multiple data entry tasks open at the same time. Multitasking between data entry tasks is a common occurrence, but this strategy can be prone to errors. The aim of the current follow-up study is to see if differences in IAC between information sources makes people more likely to interleave between data entry tasks.

The following question will be addressed:

What is the influence of IAC on interleaving between data entry tasks?

4.3.2 Method

Participants

Fourty-two participants (32 female), ranging from 18-46 years (M = 25, SD= 8) took part in the experiment. They were recruited from a university subject pool and received £4 for their participation.

Materials

The task was similar to the one used in Study 4 but differed in one aspect. Instead of filling in one data entry form per trial, participants had to complete two forms per trial, which were shown on two different pages (see Figure 4.11.Each data entry sheet contained one expense, and participants completed the trial by entering the amount and account code for each sheet. The aim of this follow-up study was to investigate if differences in IAC of the two sources makes people more likely to interleave between two separate data entry tasks.

Design

The experiment was a between-participants design with the presence of a delay as the independent variable. As in Study 4, in the Control condition there were no delays in opening the pages. In the High-Amount condition, there was a delay in opening the page with amounts. In the High-Account condition, there was a delay in opening the page with account codes. The main dependent variable was whether participants interleaved between sheets or not: did participants enter the data items in sequential order, or did they interleave between the two sheets? If participants entered the amount and account code of one sheet before entering the other sheet, this was considered a sequential order. If participants entered amounts of each sheet first, followed by entering the account codes or vice versa, this was considered interleaving. All key presses were recorded to determine in which order data was entered. Page switches were recorded to capture when and how often a participant looked up the data items. Other dependent variables were trial completion time, data entry error rate, and type of errors.

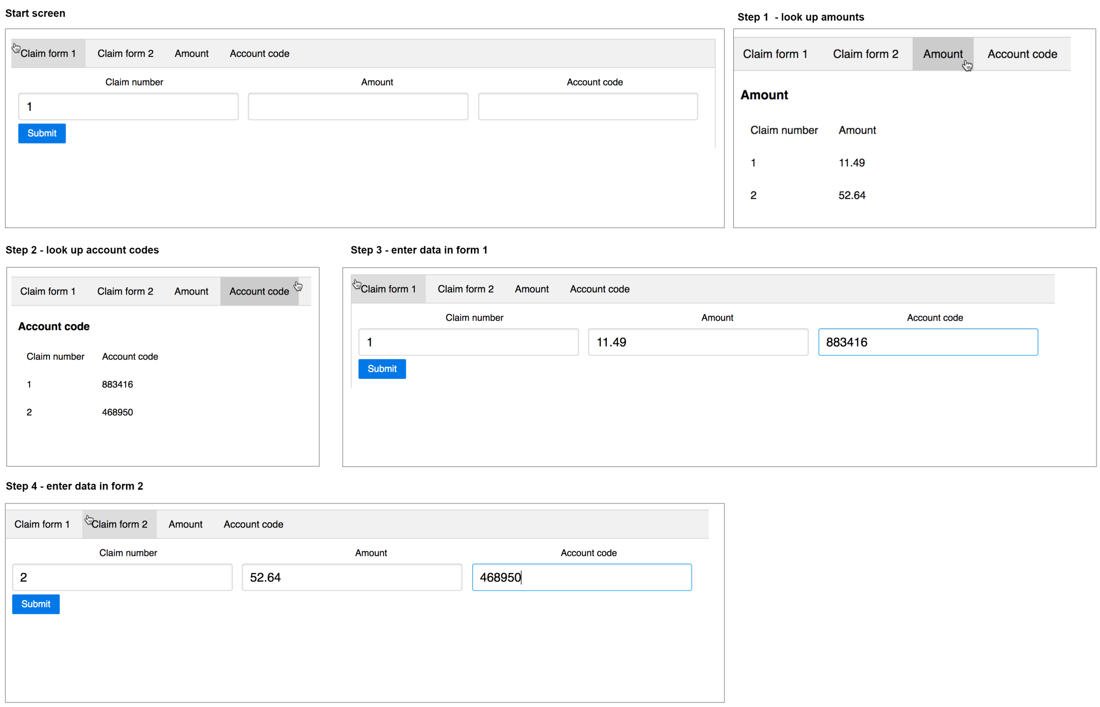


Figure 4.11: Participants had to enter two data entry forms per trial, each containing two items. Each trial started by showing the first data entry form. As in Study 4, the data items for both forms were retrieved from a separate Amounts page (Step 1) and an Accounts page (Step 2). Participants had to enter the items for the first form (Step 3) and second form (Step 4) before submitting the data entries and moving on to the next trial.

Condition	Interleaving rate	Number of visits	Duration of visits (s)	Error rate	Trial completion time (s)	Trial completion time (excl lockouts) (s)
Control	30.48% (37.71%)	4.5 (0.58)	2.00 (0.68)	3.88% (4.13%)	27.39 (3.49)	27.39 (3.49)
High-Amount	83.77% (21.59%)	4.80 (1.85)	2.61 (0.85)	7.54% (4.33%)	33.11 (8.16)	28.71 (6.22)
High-Account	73.41% (32.13%)	4.35 (1.28)	2.25 (0.67)	8.42% (9.07%)	33.83 (6.08)	29.47 (5.94)

Table 4.4: The means (and standard deviations) of all dependent measures for each condition. The rates are calculated by dividing the number of occurrences to the number of opportunities, e.g. an interleaving rate of 50 percent means participants interleaved on 50 percent of trials.

Procedure

The experimental setup was similar to Study 4. For each experimental trial, participants had to enter four data items: they had to complete two forms with two entries each, an account code and an amount. For each experimental trial, participants had to enter four data items, two for each sheet. It was explained that they could use any strategy they wanted, but that it was important to complete both sheets before continuing to the next trial. Participants first completed two practice trials to familiarise themselves with the task, and data from the practice trials were excluded from the analysis. The experiment took approximately 30 minutes.

4.3.3 Results

Three participants were removed from the data due to extreme values on performance measures. P28 and P23 made at least one error on every trial. They made 118 and 153 errors out of 200 error opportunities, respectively. P26's session was terminated before the end had been reached, as 45 minutes had passed. This participant spent on average 65 seconds per trial, which is twice as long as the mean trial time of other participants.

These three participants were considered outliers and removed from the data. Data of the remaining 39 participants was taken into the data analysis.

Table 4.4 shows a summary of the results of all three conditions for the dependent variables. Kruskal-Wallis tests were carried out to test if there were significant differences between the conditions.

Interleaving strategies

A trial was labelled as 'interleaving' if the participant started entering one data entry sheet, but interleaved to entering items on the other sheet before completing the first one. The interleaving rate for each condition was calculated by dividing the number of trials where people interleaved by the number of total trials.

The boxplots in Figure 4 show the variability of interleaving rates across conditions. Participants interleaved most often between data entry sheets in the High-Account (M = 73.4%, SD = 32.1%) and High Amount (M = 83.8%, SD = 21.6%) conditions compared to the Control (M = 30.5%, SD = 37.7%) condition, $\chi^2(2) = 11.13$, p = 0.004. A post-hoc comparison showed there was a differnece between the Control and the High-Amount (p<.01) and High-Account (p =

0.01) conditions, and no difference between the High-Account condition and the High-Amount (W = 22, p = 0.4) conditions.

As can be seen in Figure x, which shows the distribution of interleaving rates, all participants in the High-IAC conditions interleaved on at least a part of the trials. This is illustrated by the left side of the graph: the lines of the High-IAC conditions have a frequency of 0 participants at an interleaving rate of 0%. The Control condition line has no obvious peak, indicating that interleaving rates in this condition were evenly distributed: participants interleaved on zero, a portion, as well as all of the trials.

As in Study 1, participants made on average four visits per trial, i.e. one visit per data entry. There was no difference in the number of visits, $\chi^2(2) = 1.59$, p = 0.5. Participants made significantly shorter visits in the Control (M = 2.00s, SD = 0.68s) condition compared to the High-Account condition (M = 2.25s, SD = 0.67s) compared to the High-Amount (M = 2.61s, SD = 0.85s) and $\chi^2(2) = 6.14$, p= 0.04. Post-hoc comparisons found a significant difference between the High-Amount and the Control (p=.02) conditions, but not between High-Account and Control conditions (p = 0.2) or the High-Account and the High-Amount (p = 0.2).

Errors and trial completion time

There were 200 data entries, so in total there were 200 opportunities for a participant to make a data entry error. The error rates were calculated as the number of errors divided by the number of entries. There was a marginal though not significant effect of IAC on error rate, $\chi^2(2) = 5.37$, p = 0.06. The mean error rate was marginally higher in the High-Account condition (M= 8.42%, SD=9.08%) compared with the High-Amount (M=7.54%, SD=4.33%) and Control (M=3.88%, SD=4.13%) conditions. When comparing the actual completion time including lockouts, participants were significantly faster in the Control condition (M=27.39, SD = 3.49s) than the High-Account (M = 33.83s, SD = 6.08s) or High-Amount (M = 33.11s, SD = 8.16s) conditions, $\chi^2(2) = 8.52$, p= 0.01. With the lockout times removed, the difference is no longer significant, $\chi^2(2) = 1.61$, p = 0.4.

The type of errors can be seen in Figure. The most common error type was when a data entry was skipped: this happened 243 times. Table 1 shows the number of skipped errors for each condition. It can be seen that in the Control condition this type of error occurred 16 times. The error happened more frequently in the High-IAC conditions: in the High-Account condition it happened 114 times, and in the High-Amount condition it happened 116 times. Typing the correct number but in the wrong field happened 78 times. This happened 18 times in the Control condition, 14 times in the High-Account and 46 times in the High-Amount condition. When comparing across conditions, these types of errors happened on a significantly higher proportion of data entries in the High-Account (M = 4.58%, SD = 3.6%) and High-Amount (M=6.54%, SD=5.01%) compared with the Control condition (M = 1.23%, SD = 1.82%), $\chi^2(2) = 11.29$, p = 0.004. A post-hoc comparison showed there was a difference between the Control and the High-Amount (p<.01) and High-Account (p = 0.01) conditions, and no difference between the High-Account condition and the High-Amount (W = 22, p = 0.4) conditions.

4.3.4 Discussion

The aim of this study was to see if an increase in IAC makes people interleave more between data entry tasks. In contrast with Back et al. (2012), who found that an increase in IAC made people less likely to interleave between two data entry tasks, participants in the current experiment interleaved more as IAC increased.

This may be due to the presentation of the information. In Back et al. (2012)'s study, people had to retrieve all information for both data entry tasks from one sheet. If the sheet was nearby, participants read one item at a time, and interleaved between tasks on 59% of the trials. As the cost to access this source increased, they chunked the data items associated with one task, and then after completing this task, returned to the source to chunk data items for the second task. However, in many situations, such as the office setting of Chapter 3, information is not in one location, but different information sources have to be consulted for different types of information. For an expenses task, amounts and account codes are not on one sheet, but people have to consult one spreadsheet for account codes, and another source to retrieve the amounts. The current study looked at people's interleaving behaviour when retrieving items from multiple sources. If there were no delays in accessing these sources, participants completed one task at a time on 59% of the trials. If there was a delay in accessing either one of these sources, people tried to enter all information from this source after one visit, so they did not have to open it again. They chunked the data items associated with one source, rather than task. They first entered either Amount1 for Form1, Amount2 for Form2, and then the accounts, or first entered Account1 for Form1, Account2 for Form2, and then the amounts.

Whereas in Study 4 people became more accurate by chunking data items according to IAC, there no longer was a difference in errors in the current study. Chunking by data items in this set-up meant people interleaved between tasks and started a second task before completing the first task, a strategy which can be prone to errors (). People may forget steps, or enter correct information in the wrong fields.

It can be argued that the design of the materials encouraged participants to always group per source, regardless of the condition. However, in the Control condition there was an almost even distribution of strategies, and participants interleaved 40% of the trials. The majority of the time participants still chose to complete one task at a time.

4.3.5 Conclusion

People have to regularly switch between looking up information for a data entry task and entering it. The three studies described in this chapter showed how strategies to look up and enter information are influenced by the time cost to access information sources. It also showed that certain strategies are more accurate or efficient than others. The main effect of an increased IAC is that people try to minimise (re)visits. If the time to access a source increases, people will try to copy over more information after one visit. If they do not memorise it well, errors increase (Study 3). If information is spread across different sources and the IAC differs between these sources, people group visits and first look up and enter low IAC items, before entering high IAC items. This not only made them more efficient, but it also reduced errors (Study 4). However, if they have to manage multiple data entry tasks, this strategy means that they will interleave between tasks (Study 5).

These results are partly in line with observations from the first two studies. Whereas people would look at the physical receipts while typing it in, they would hold other items in memory and barely used tools to offload memory. They would first enter all items on the physical receipts. For digital information however, they would look it up as they needed it, even if IAC differed between these sources, and it could sometimes take a while before they had retrieved the information.

It seems that it is better to be able to reduce IAC and have task information ready at hand, so people do not need to switch back and forth to a source that takes time to access. There are solutions, such as increased screen space, multiple screens, or having a physical copy and placing it nearby. It was interesting to see in the first two studies that people did decrease IAC for physical items, but not for digital ones. People had a second screen, but used their primary screen to look up information because they perceived it as quicker. In this case, the cost to decrease IAC by placing information on a second screen, outweighed the cost to look it up, hold it in memory, and go back to the primary screen. However, they often did not know the associated time cost to access it, so could be away from the screen for a longer time than anticipated.

There is a need to better support people in decreasing IAC without tasking them with the added responsibility of re-arranging different tasks, information sources, devices and screens. They should be able to do their job and have task information at hand more seamlessly.

Chapter 5

Design considerations

Chapter outline

Based on both findings from my studies and previous literature on information search and interruption management, this chapter explores a range of design possibilities to better support people in managing interruptions to collect information.

5.1 Introduction

The previous studies reported in this thesis have shown that when looking up information for data entry work, people adopt different strategies. Study 2 showed that people group paper sources and collect this beforehand, but digital information is looked up when needed. Affordances of digital sources are more hidden than affordances of physical artefacts (Sellen & Harper, 2003), and it is less visible where to get information from and how long it will take to find information. As a result, people interrupt their data entry task as soon as they need digital information. Study 4 and 5 showed that if people learn how long it will take them to access information, they adapt their planning strategies. If participants knew it took them a long time to access a source, they postponed looking it up and entered other information first. An issue highlighted in Study 2 is that people often do not know where to get digital information from, and do not know if it will take them a long time until they have already interrupted themselves. Furthermore, whereas in Study 4 and 5 people always needed to use the same information from the same sources, participants in Study 1 and 2 did not always know they needed information until they had started a task. As soon as a need emerged, they addressed this need immediately to not have to hold it in memory. Interruptions are disruptive for data entry in a number of ways. It takes time for people to resume the task, they may have forgotten where they were and enter information in the wrong fields (Brumby et al., 2013; Monk et al., 2008)..

A large body of work has looked at how people organise, manage and re-find information, in order to design appropriate information management tools (Dumais et al., 2003; Trullemans, Sanctorum, & Signer, 2016). However, most of these studies tended to focus on finding information as a main task, and not as a subpart of other work and how people schedule interruptions from work to look up information. Furthermore, the tasks studied to evaluate these tools usually involved organising documents that had to be used for a longer time. The type of data entry task central in this thesis is characterised by rapidly going in and out of many different types of information sources. The current chapter reviews prior relevant work on information management and search tools, and explores a set of design possibilities on how people can be supported in managing interruptions to look up information.

5.2 Related work

Switching between documents and applications to look up information is common for many activities beyond data entry work. For example, people need documents when writing a paper, or use emails, calendars and written documents to plan a project. Prior research has explored tools to support fragmented work has focused on information management, information search, and integration of information from multiple sources.

5.2.1 Task and information management

To make it easier to re-find documents for a particular activity, some systems have looked at grouping windows and documents. For example, TagFS (source) allows users to tag documents and define groups. GroupBar (Smith et al., 2003) makes it possible to group windows needed for a task in the task bar. This can be particularly useful when resuming an interrupted task: the user can see which documents were used before leaving the task. These tools offer the user flexibility in organising information sources in different ways, but come with a number of limitations. First, it assumes the user knows in advance what information is needed for which purpose. While some information needs are known in advance of the data entry task, it regularly occurs the user needs unexpected information. Second, categorising documents can be time-consuming, and people are often not willing to invest time doing so (source). Especially for data entry tasks where documents are only briefly needed, people may not make the effort to group information. Lastly, studies have shown that when people do make the effort to organise documents into groups, they often use inconsistent labels, making it difficult to re-find information later (Jones, Jiranida Phuwanartnurak, Gill, & Bruce, 2005).

5.2.2 Information search

In addition to information management, other studies have focused on supporting information search. An issue with looking for information is that information can be scattered across applications, and users have to go in and out of these separately to search and find what they are looking for. To support re-finding information across different applications, Dumais et al. (2003) presented a tool called Stuff I've Seen. Users were presented with a unified search interface which they could use to search through information they had already seen before, such as emails and web pages. A user study, where participants installed the tool on their computer and used it for two weeks, showed that users used the search tools of individual applications less frequently and used Stuff I've Seen instead. The focus of the tool was to improve search, rather than the scheduling or reducing of interruptions from work to search. The user still had to switch from their main task environment to a separate tool, and create a search query or use filters to view relevant search results. People do not always know what to search for, as demonstrated by both previous literature (source) and Study 2 of this thesis. Furthermore, for familiar documents, the preferred way of navigation is often browsing, rather than searching (Bergman, Beyth-Marom, Nachmias, Gradovitch & Whittaker, 2008). Whereas Stuff I've Seen only supported searching for digital information, PimVis was developed by Trullemans, Sanctorum, & Signer (2016) to allow search across both paper and digital sources. A graphical user interface presented a visualisation of documents, grouped according to the context in which they are relevant. Bookcases and filing cabinets were augmented with LEDs, which would light up if users selected a document in PimVis that was contained in these physical locations. By opening a document in PimVis, the user could see documents related to this document. PimVis was evaluated using the task of finding documents for writing a paper. As PimVis was a standalone application, users had to switch away from their current application, such as their text editor, and open the document in PimVis to view its related documents. Participants reflected that PimVis would be useful for archived documents. For so-called 'hot' documents, which are needed for short-term tasks in the moment, they would value seeing related documents in the environment they are currently already working in, rather than having to go to a separate application. In the user study, the grouping of documents as well as augmentation of physical artefacts were set up by the researchers. The primary aim was to see whether participants could easily navigate through PimVis. They were given tasks to find specific documents, such as 'You want to read the paper called X, which is related to the paper called Y'. In practice, the user would have to pre-define in which contexts documents were to be used and how they were related to other documents, which has the same drawbacks as categorising and labelling documents as discussed above.

5.2.3 Interruptions and delayed intentions

One reason why people are not always able to organise information efficiently is because they may not know they need information until they have started a task. In Study 2, participants disrupted their data entry work as soon as they realised they needed certain information. Prior work on self-interruptions found that office workers often start new tasks before completing old ones (Czerwinski, Horvitz, & Wilhite, 2004; Jin & Dabbish, 2009). If people are able to keep track of tasks they need to perform, it may help them in deferring these tasks until a more convenient moment, rather than addressing them as they realise they need to be done (Jin & Dabbish, 2009). For example, an interruption between subtasks is less disruptive than an interruption in the middle of a subtask (Bondarenko, 2010; Gould, 2014). Gilbert (2015) looked at people's off-loading behaviour in both an experimental and naturalistic setting. Participants had to remember to perform an action later, and had the option to offload this intention or to keep it in memory. In both settings, a majority of participants offloaded these intentions when they had the option, and this significantly improved their performance. Additionally, in Study 3 of this thesis, where participants had to remember which blocks to drag to which location, a selection of participants placed blocks nearby what they though the correct location was, to not have to remember its location, and as a reminder to place them there later. These findings suggest that if people have to memorise which information to retrieve, they may benefit from options to offload these information needs, and are able to effectively defer information subtasks until a convenient moment in the main data entry task.

5.2.4 Documents at hand

Bondarenko & Janssen (2005) compared how information workers store paper and digital sources. One user need they found was that documents should be embedded in task-related context information, as it helps to resume a task after an interruption. In addition, in Bi & Balakrishnan's (2009) study on large and multiple display use, office workers felt more focused on the task and immersed in their work when surrounded by task-relevant documents. A limitation of most tools discussed so far is that the task window and information window are separate, and users need to switch between these. Microsoft Office's new feature TAP instead is a built-in add-on, which allows users to place relevant documents in a task pane next to their working document. The aim of the feature is to keep focus on document creation, rather than looking up information. The feature is presented as a task pane within a document, such as a text document or email, and contains an overview of documents that may be relevant to the current document. The task pane initially shows files that are most frequently used. If the pane does not show the documents that the user is looking for, there is also a search option within the task pane. The feature can reduce interruptions from the task interface. However, the TAP feature is application-specific: the user is only able to include other Office documents, and not information sources such as websites and databases. Furthermore, it is mainly focused on re-using content from archived documents, and assumes the user knows which documents to re-use. It may be less suitable for situations where people do not know where to get information from.

5.2.5 Type of activity

An important difference between previous work on information management and the studies in this thesis is the nature of the activity studied. Bondarenko & Janssen (2005) distinguish between two types of activities that information workers engage in: research activities and administrative activities. For an administrative activity, users go in and out of a variety of documents rapidly. For research activities, users need a smaller variety of documents, but these are needed for a long time. The tasks studied in most information management studies were more similar to research activities: participants had to read information to improve their understanding of a legal case (Cangiano & Hollan, 2009), or they needed to have the information for a longer time during a task (O'Hara, Taylor, Newman, & Sellen, 2002). A data entry task however is more similar to an administrative task. This distinction in activities is important, as it influences people's information collection strategies (Bondarenko & Janssen, 2005), and whether design considerations from previous work are applicable. Participants may not want to spend effort to organise information, such as grouping them and indexing them, if they only need the documents briefly. On the other hand, having contextual information nearby can be beneficial for both types of tasks, as it minimises interruptions and holding items in memory.

5.2.6 Time management applications

In order to improve focus and mitigate self-interruptions, Kim, Cho and Lee (2017) developed an intervention that allowed people to temporarily block specific sources that they considered distracting, such as email, IM applications and social media. However often these sources then needed to be accessed after all for the task they were working on. Other commercial applications do not block sources but instead provide users an overview of their computer activities, to reflect how much time they spend in total on tasks, and certain sources ("ManicTime," 2018, "RescueTime," 2018). However, as these tools provide information of past usage, it is often not clear to users what they have to do with the data (Collins, Cox, Bird, & Cornish-Tresstail, 2014), and there is little evidence of their effectiveness in improving focus (Whittaker, Hollis, & Guydish, 2016). Gould, Cox and Brumby (2016) looked at switching behaviour during online crowdsourcing work, and found that an intervention during work that encouraged people to stay focused after people had interrupted reduced number of switches to unrelated tasks. Recognising that switches occur as part of the task, we consider whether the duration of switches can be reduced by giving people real-time feedback on how long they switch away for during a data entry task. This is important to consider, because the longer people interrupt, the more disruptive it is (Monk, Trafton, & Boehm-Davis, 2008), and the harder it is to resume a task (Altmann, Trafton, & Hambrick, 2017).

5.2.7 Interruptions

5.2.8 Summary

Previous work on information search has looked at improving search across applications and media, but provides limited support for users on when to interrupt their work to conduct these searches. Prior work on information management has found that having contextual information at hand can reduce interruptions and helps users to be focused on their work. However, many of these tools require the user to organise, file and tag documents beforehand, and are based on the assumption that users know where to get information from. If people are able to off-load intentions to look up information, they may efficiently schedule when to interrupt their task and collect information.

5.3 DESIGN CONSIDERATIONS

Based on both findings from the literature review and findings from my studies, the following requirements for information tools for data entry work can be defined: 1. Users should be able to group information for a task. 2. (Users should be able to search for different types of information sources.) 3. Users should be able to keep track of information they need. 4. Users should be able to off-load intentions for subtasks of looking up information. 5. The information interface should be embedded within the main task interface. Three design alternatives are proposed below. For each design feature, Table 2 summarises which findings it builds on, and whether these findings are derived from previous research and/or my studies.

5.4 Design alternatives

Chapter 6

Evaluating a time notification on interruptions

Chapter outline

This chapter describes two studies that, given varying IACs, evaluate whether giving people feedback on the duration of their switches can influence people's switching strategies and data entry performance. A browser notification was developed which showed people how long they switch away for on average. Study 6 evaluates the notification with an experimental task, to see if time information influences the strategies people adopt, and can make people adopt more accurate and faster. Study 7 evaluates the notification in the office setting with workers doing data entry work, to ascertain how appropriate the proposed recommendations are for a naturalistic task for which they would be used.

Together these studies intend to show that time information makes people switch less between entering and looking up information, which makes them faster to complete the data entry task overall and can reduce errors.

6.1 Chapter introduction

Chapter 3 presented two qualitative studies that investigated how office workers perform data entry work in office settings. It revealed that self-interruptions to collect information is a major aspect of their work. Even though data workers tried to collect information carefully before starting a task, they often did not realise they needed certain information until starting the task. Whereas they deferred retrieving physical sources, for digital sources they self-interrupted their task immediately to retrieve additional information. The hypothesis was made that differences in expected time costs influenced people's decisions on when to switch and interrupt themselves.

Chapter 4 presented findings from three controlled experiments that supported the hypothesis: when time costs to access digital sources are consistent, and people learn and anticipate the time it will take them to retrieve information, they will look up and enter items that take the least time first, and postpone getting information that takes time to look up. An issue is that, outside of a controlled setting, people often do not know how long an interruption to look up digital information may take, and whether they should switch immediately or later. As was observed in Study 2, digital interruptions often took far longer than intended because people had to search through large and multiple documents, and could get distracted by irrelevant information. Because they did not know the time they spent on task interruptions, it was difficult to manage these self-interruptions.

The studies presented in this chapter aim to investigate whether giving people feedback on the time spent on task interruptions has any effect on people's self-interruptions. In particular, the purpose is to see whether time information reduces the number and duration of interruptions, and as a result can improve data entry performance. Study 6 used an experimental data entry task to measure if a notification showing the average duration on people's window switches had an effect on number and duration of their switches, and data entry speed and accuracy. Study 7 evaluated the feedback with data workers doing expenses work, to evaluate if the notification would be suitable for an applied task.

6.2 Study 6: Looking up information in email during an online experiment

6.2.1 Introduction

This study aimed to investigate whether an intervention showing people how long they switch on average has an effect on the duration and number of switches during a data entry task. An online experiment was conducted where participants had to complete a data entry task. Participants had to enter numeric codes into a form, which they had to retrieve from a message sent to their personal email. The information was presented as a message in participants' email inboxes, as email is an integral part of data entry work but known to be a source of distraction, and people often spend more time on it than originally intended (Hanrahan & Pérez-Qu, 2015; Mark, Iqbal, Czerwinski, Johns, & Sano, 2016). It was therefore expected to have a distracting effect during the switches to look up information. Half of the participants received feedback on the average length of their switches through a browser notification. The research questions that this study aims to address are whether feedback on interruption length:

- reduces the number of switches?
- reduces the duration of switches?
- makes people faster in data entry?
- makes people more accurate in data entry?

6.2.2 Method

Participants

Fourty-seven participants (30 women) took part in the online experiment. Ages ranged from 20 to 63 (M = 29.3 years, SD = 9.1 years). The participants were recruited via university email lists, social media and online platforms to advertise academic studies, and participation was voluntary. Participants were alternately allocated to the control or experimental condition.

Design

The study used a between-participants design with one independent variable, a notification. In the control condition, participants did not receive a notification, but switches away from the

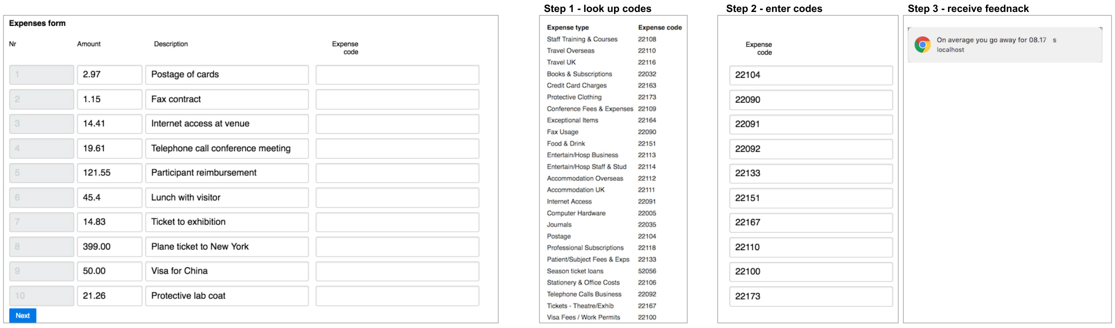


Figure 6.1: The data entry task as shown in the browser. Participants had to look up codes from their email (Step 1) and enter this into a sheet (Step 2). After every trial, the notification condition received time information (Step 3)

data entry window were recorded. In the notification condition, participants were shown a notification every time they completed a trial. This notification showed how long on average they were away for when switching away from the window, before returning to the task. The purpose of this notification was to see if the number and duration of switches could be reduced by giving participants feedback on the time spent of on switches. Dependent variables were number and duration of switches away from the data entry interface, trial completion time, and data entry errors. Switching behaviour was recorded using JavaScript's blur and focus events. These were triggered whenever a participant switched away from the data entry window, whether to their email inbox or to a different window or application.

Materials

The task used was based on a common routine data entry task from Study 1 and 2 involving processing expenses. Participants were presented with an online sheet containing a set of ten 'expenses' (see Figure 1). They had to complete each row by entering the correct expense code for the expense. They retrieved this code by looking it up in a table of 25 expense categories which each had a corresponding 5-digit expense code, shown in Figure 2. Participants had to determine which category an expense belonged to, look up the code of this category and enter it in the row of the expense. We used expense categories and codes that are currently used by a public university to process expenses.

In the example of Figure 1, the expense in the top row belongs to the category 'Postage' and the participant would have to copy the code 22104 from the expense table into the empty cell of the top row. A code did not occur more than once in a trial. The codes within a trial could be entered in any order. Once the codes of the ten expenses had been entered, participants clicked the Next button to go to the next trial and the sheet was filled with ten new expenses. Participants were not alerted to any mistakes and once they had pressed 'Next', they could not return to the previous trial to correct any errors. Participants had to complete one practice trial, and five experimental trials. The purpose of the practice trial was for the participant to get familiar with the task, and the recorded data from this trial was excluded from the analysis.

The experiment was conducted in a web browser. In addition to the main task, we implemented a browser notification that appeared when participants in the notification condition switched away from the data entry window (see Figure 3). Every time participants switched, a notification appeared at the right-hand corner of their screen that told participants how long on average they go away for when they switch. The notification stayed visible for several seconds as set by default by the browser, or participants could dismiss the notification themselves by clicking on it.

Procedure

The study was advertised online with a brief description and a website link to sign up. Participants signed up for the experiment by entering their email address, and were sent an email with the table of expense categories and expense codes. The email also included instructions with a new link where the study was available. Participants were asked to complete the task on a desktop or laptop computer and open the experiment in Google Chrome, Firefox or Safari. Participants were not informed beforehand which condition they had been allocated to, and were told the purpose of the study was to understand how people perform data entry tasks. Participants in the notification condition were informed that they would receive notifications during the experiment. Participants first read an online consent form on the website, and were not able to continue to the experiment until they had agreed to the consent form. Participants in the notification condition received an additional dialog box to enable notifications in their browser, and had to click 'OK' to continue. Participants were instructed to have both their email and data entry window open on the same device, and to keep both windows maximised at all time, to ensure they had to switch back and forth between the two windows. Participants who made no recorded switches would be excluded from the dataset. After completing all experimental trials, participants were shown a page of debriefing information, explaining the purpose of the study. An email address was included as a point of contact if participants had any further questions. Participants took between 10 and 20 minutes to complete the experiment.

Pilot study

6.2.3 Results

Table 6.2 summarises the results of the conditions in terms of the four dependent variables. The number of switches, length of switches and the error rate were not normally distributed, so non-parametric Mann-Whitney tests were used to analyse effects of a notification on these dependent variables. A Shapiro-Wilk test suggested that the trial completion times were normally distributed, $W = 0.97$, $p = 0.22$, so an independent t-test was used to analyse the effect on trial times.

Number and duration of switches

Figure 6.3 shows the variability of duration of switches for the two conditions. Results show that switches were significantly shorter among participants who had a notification ($M = 4.76$s, $SD = 1.65$s) than among those without a notification (M=7.13s, SD=3.05s), $U = 406$, $p < 0.01$, r=.44. Participants switched once for every code entered (i.e., ten times per trial), and there was no significant difference in number of switches between conditions, $U = 243$, $p = 0.6$.

The distribution of switching durations was positively skewed with a long tail: switches up to 11 seconds made up 90% of switches, but the longest switch was greater than seven minutes. The distribution of the top 10 % of longest switches is illustrated in Figure 6.2. Table **??** shows the count of extremely long switches per condition.

Task performance

Error rates were calculated by dividing the number of data entry errors divided by error opportunities. The error rates were significantly lower for participants with a notification ($M = 2\%$, $SD = 2\%$) compared to participants who had no notification ($M = 5\%$, $SD = 5\%$), $U = 403$, $p < .01$, $r = .44$. Participants with a notification were also faster in completing trials ($M = 107.61$s, $SD =$

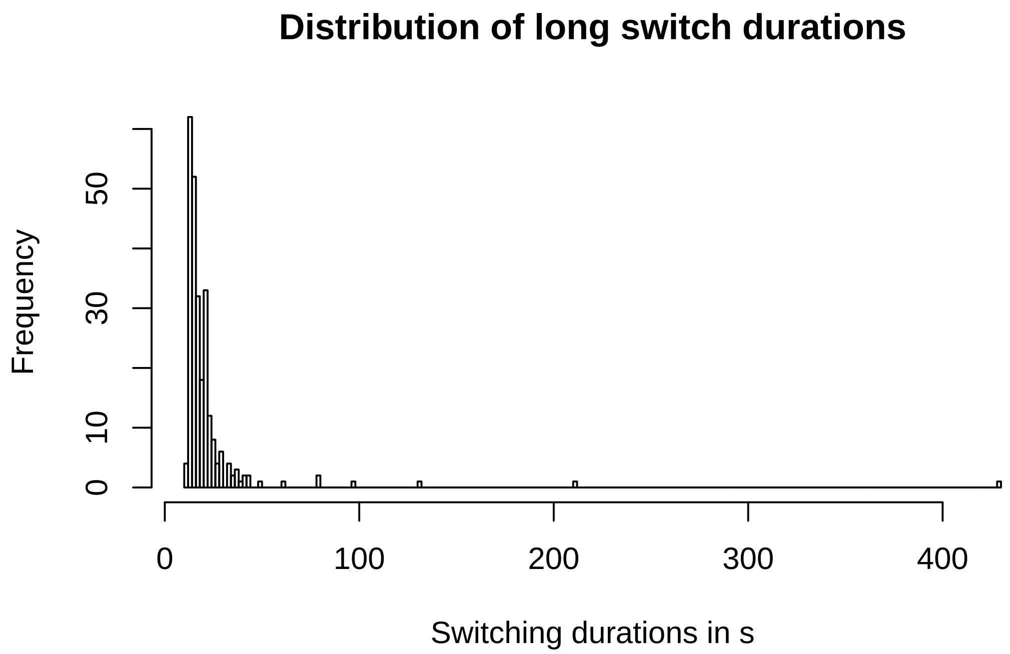


Figure 6.2: Distribution of the top 10% longest window switches, that is switches that were longer than 11.82 seconds.

Table 6.1: Total number of switches at different percentiles for each condition.

Percentile	Duration (s)	Control	Notification
90th	11.82	168	85
95th	16.74	89	38
96th	18.27	75	27
97th	20.30	53	23
98th	22.18	32	19
99th	29.13	13	13

Table 6.2: Means and standard deviations of dependent variables for each condition.

Condition	Number of switches	Duration of switches	Error rate	Trial completion time (s)
Control	10.65 (1.67)	7.13 (3.05)	5% (5%)	126.27 (32.61)
Notification	10.61 (1.83)	4.76 (1.65)	2% (2%)	107.61 (31.15)

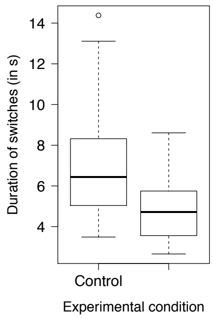


Figure 6.3: Boxplot of duration of switches away from the data entry interface in each condition.

31.15s) compared to participants without a notification (*M* = 126.27s, *SD* = 32.61), $t(45) = 1.98$, $p < .05$, $d = .59$.

Interkey intervals

The primary measure to analyse switching behaviour were focus and blur events. These measures include any switch from the task window to another computer window, but task switches outside the device, with the task window still in focus, were not captured. Therefore, inter-keystroke interval (IKI) data was analysed to look for large intervals. The IKI data presented here does not make a distinction between moments when participants were inside or outside the task window, and longer intervals may have also been moments were participants had briefly paused for thought. However, extremely large intervals between two keystrokes might point to moments where participants had switched to doing something else.

Figure 6.4 shows the distribution of IKIs. Table **??** shows the count of extremely long IKIs per condition.

Table 6.3: Total number of IKIs at different percentiles for each condition.

Percentile	Interval (s)	Control	Notification
90th	6.21	1143	451
95th	9.82	572	226
96th	11.14	453	185
97th	12.94	335	144
98th	16.41	214	105
99th	23.02	103	57

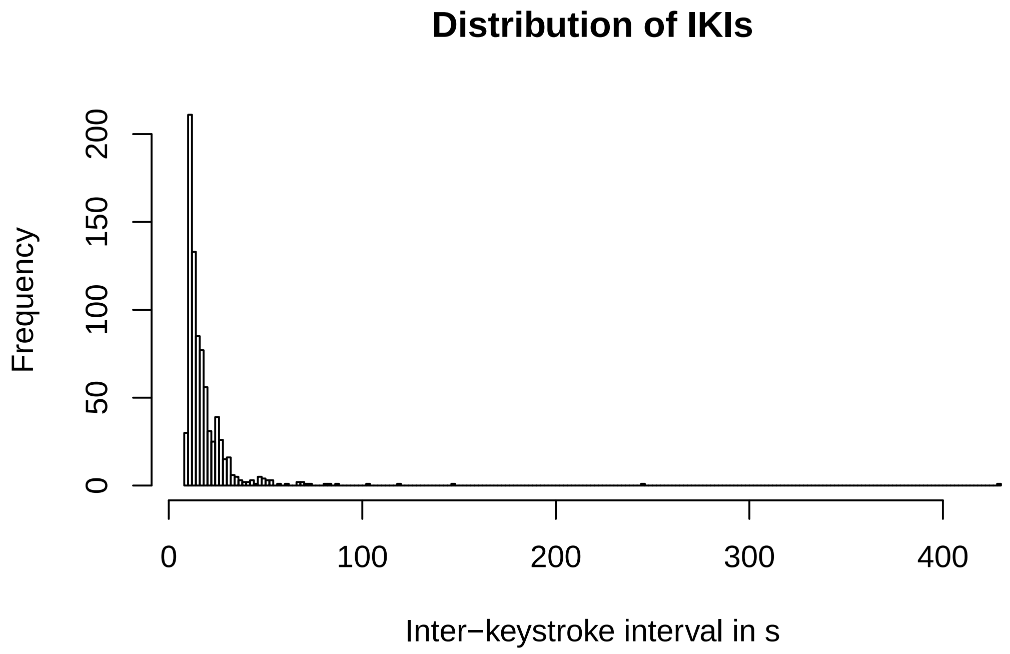


Figure 6.4: Distribution of the top 10% longest IKIs, that is IKIs that were longer than 6.21 seconds.

6.2.4 Discussion

The aim of this study was to see whether showing people how long they switch on average reduces the number and length of their switches. The results show that people can benefit from receiving feedback on the length of their switches: participants made shorter switches, were faster to complete the task, and made fewer errors. These findings suggest that shorter switches can lead to better task performance, and are in line with previous studies connecting the duration of an interruption to its disruptiveness (Altmann, Trafton, & Hambrick, 2017; Monk et al., 2008).

Nevertheless, as even short interruptions can have a negative effect on performance (Altmann, Trafton, & Hambrick, 2014), it was also measured if the number of switches were reduced. Interestingly, feedback on switching duration did not reduce the number of switches as in prior work (Gould, Cox, & Brumby, 2016). This could be explained by the moment in the task that people received feedback. In Gould et al.'s study, feedback appeared after every switch. Participants may have tried to reduce switches, either because they were more aware of every switch or because they wanted to avoid the message. In contrast to our study, their participants were not supposed to switch, so the number of switches was lower. In the current study participants were switching more often as they had to as part of the task: on average, they switched once for every data entry (i.e., ten times per trial). Giving notifications at every switch would have had the risk of overexposing participants to notifications and limiting its usefulness (Cutrell, Czerwinski, & Horvitz, 2001; Whittaker, Hollis, & Guydish, 2016). Therefore, feedback was only given after every trial. Future data entry studies that require fewer switches are needed to see if a notification upon every switch can reduce both the number and length of switches. Moreover, because the notification only showed information regarding the duration of switches, participants may have focused on reducing the duration, rather than number of switches.

The current study used focus and blur events to analyse switching behaviour. This meant that task switches outside the device, with the task window still in focus, were not captured. Possibly participants learnt to not interrupt themselves when they were away from this window, but after they had returned to the window. Without an accurate estimate of how long participants should take to complete the task, it is difficult to determine moments at which participants were away from their computer (Rzeszotarski, Chi, Paritosh, & Dai, 2013). Using other techniques, such as prompts at random intervals to confirm people are still working on the task, may be able to give a further insight whether our intervention changes overall self-interruption behaviour.

Most studies on self-interruptions introduced an artificial distraction, such as chat messages, to measure when, how long, and how often people self-interrupt to attend to this distracting task (Katidioti & Taatgen, 2013; Salvucci & Bogunovich, 2010). The current study makes a methodological contribution by using participants' own personal email inbox, based on the assumption that email provides a source of distraction (Hanrahan & Pérez-Qu, 2015; Mark et al., 2016). However, in the current study, participants only needed to find and open an email once. Once they had this email opened, they did not have to re-find it in their inbox for the remainder of the experiment, and may have had this email maximised on their screen, hiding incoming messages. In practice however, people have to first find the email in their inbox, which can partly contribute to the distraction. Our study has already shown an effect on behaviour by switching to an email inbox. It is expected there might be a higher potential for distraction if people have to also find the correct email in their inbox.

Bridge to next study

The results of this experiment indicate that showing people how long they switch on average reduces the duration of switches and can improve people's task performance. The work makes a contribution to our understanding of switching behaviour for routine data entry tasks to distracting, but task-relevant, applications such as email. The results also suggest ways in which tendencies to attend to distractions might be mitigated, and can provide a useful pointer for the design of productivity interventions to improve focus. In the current study, an experimental task was used in order to measure task performance.

The study did not find any effect on the number of switches. However, as the number of switches indicate, participants presumably only switched to look up information, and did not switch to other tasks, interruptions or distractions. When people are doing their own data entry work, they may not have to switch as often as in the current experiment, but on the other hand they may interrupt themselves more to attend to other tasks and interruptions. To evaluate whether the positive effect of time feedback on people's switching behaviour can extend to naturalistic tasks, Study 7 tested the notification with office workers doing their own data entry work,

6.3 Study 7: Looking up information for expenses in an office setting

6.3.1 Introduction

In order to understand whether the notification would have the same effect on a naturalistic data entry task, Study 6 was followed up with a field study testing the notification with data entry workers doing expenses work. To measure self-interruption behaviour during their work, participants were asked to install a free trial version of ManicTime [1] for two weeks. ManicTime is a time tracking software, which tracks application and web page usage. In addition, half

[1] https://www.manictime.com

of the participants were asked to install a browser extension, and use it when they are processing expenses. Every time participants switched away from the browser window in which they did their expenses work, the extension showed a notification similar to the notification used in Study 6. The purpose of the study was to see whether a notification had an effect on self-interruption behaviour. To get a quantitative measure of self-interruption behaviour, ManicTime data was used to derive number and duration of window switches during expenses work. In addition, participants were interviewed to explore whether and how the use of both the extension and ManicTime led to any conscious changes in their behaviour.

The study aims to address the following research question: how does feedback on interruption length have an effect on people's self-interruption behaviour during expenses work in a finance office setting?

6.3.2 Focus, a browser extension

The browser notification was implemented as a Google Chrome extension, using HTML, Javascript and CSS. Participants were sent a link to download the extension, named Focus, and were given instructions to install and add it to their Google Chrome browser. To use the extension, participants had to navigate to the web page in their Google Chrome browser where they had to complete their work and click on the icon of the extension. As was found in Study 1 and 2, the majority of data entry work studied in this thesis, such as processing expenses, was done in a web browser. Upon clicking on the icon, a pop-up appeared saying that the current web page was now the main task page. Every time participants switched away from this page, they received a notification indicating how long their switches away from this page are on average. Participants could stop the notifications by refreshing or closing the page.

The main difference in notification behaviour with respect to Study 6 was that participants received the notification every time they switched away from their task window, rather than at particular moments in the task. They only received it when they switched away from their task window, but not upon any subsequent switches. Furthermore, due to security issues the extension did not store any data. Instead, the application ManicTime was used to measure people's switching behaviour.

6.3.3 Method

To study the effect of a notification on people's self-interruption behaviour, a between-groups design was used and participants were divided into a control group and experimental group. The control group was asked to install ManicTime for two weeks. They were told the purpose of the study was to understand how people in offices manage tasks, windows and applications. The experimental group additionally were asked in the second week to install Focus. They were told that this extension would give additional information on the current task they were working on. Other than this distinction, all instructions were identical between the two groups.

Four participants were unable to use Google Chrome, and therefore the extension, for their work. Therefore, these participants were part of the control group. To make the groups even, one other participant was randomly chosen to be allocated to the control group; the remaining participants were allocated to the experimental group.

Participants

Ten participants (three male) took part in the study. They were recruited using the same recruitment methods as Study 1 and 2. Invitations were sent to opt-in mailing lists of Finance departments of a university, and forwarded by contact persons and people who had already participated. None of the participants had taken part before in any of the studies reported in this thesis, but were drawn from the same population. None of the participants had used a time management application such as ManicTime before. Participants were reimbursed with a £20 Amazon voucher.

Procedure

Participants who expressed interest were sent an information sheet and consent form to read and sign. They were sent an overview of the study, instructions to install the tools, and a post-study interview was scheduled. The study was divided into three stages:

Week 1: Install ManicTime In the first week, participants were sent instructions to install ManicTime on their work computer. They were given the option to pause or stop the application from running at any time. They were told that they were free to choose if, when and how often to look at the information, but that it was important to complete at least one expenses task with the application running.

Week 2: Install Focus In the second week, participants in the experimental condition were asked to install Focus. Again, they were instructed that they were free to choose when and how often to use the extension, but that they had to use it for at least one expenses task. Even though the second week was the same for participants in the control condition, they were also asked to complete at least one expenses task in the first week, and another expenses task in the second week. This instruction was included to compare if any observed changes in switching behaviour were due to the browser extension, or if participants simply became more aware of their switching behaviour in the second week.

Week 3: Interview In the third week, participants were interviewed about how they currently manage documents, applications and tasks for their work, and asked questions on their experience of using the tools. In particular, it was discussed whether and how they used or would use the information that the tools provided, and whether they made any changes on how they went about their work. They were asked to share their ManicTime database for further analysis. Participants were offered guidance and assistance on deleting or adapting any data in their database, such as removing application and website names. Participants were still eligible to participate, if they did not wish to share their database.

The interviews took place in a closed room , lasted about 50 minutes and were audio recorded.

Pilot study

A pilot study was conducted with two participants. One participant was a colleague and was sent instructions to install Focus on his computer. The purpose of this pilot session was to see whether the installation instructions were clear and to test the extension on other people's computers. The participant had the tendency to hover over the notification to stop it from disappearing, which placed extra buttons over the end of the notification message. Therefore, the message was shortened and the most important information, namely the duration of switches, was placed at the front of the message to ensure it was still visible upon hovering.

A second pilot session was then conducted with an administrator working at the same university as the study participants, who was asked to install ManicTime and Focus on her work computer. The purpose of this session was to ensure the tools could be installed on the work computers of the university, and that the extension worked with the university finance system.

Ethical considerations

Participants were informed before undertaking the study that they would be asked to share their ManicTime database at the end of the study. However, a disclaimer was added in the invitation and instruction emails that participants were still able to take part, if they did not wish to share their database. It was made clear what data ManicTime recorded, and that Focus did not store any data. They were given instructions on how to pause the application from running and how to delete certain parts of the data, and were offered assistance to help further change the database into a state they were comfortable sharing.

Data analysis

Though ManicTime was piloted on a work computer of the university before starting the study, three participants were unable to install ManicTime on their work computer due to firewall restrictions. Furthermore, two participants opted out of using Focus as they used a browser different from Google Chrome for at least parts of their work. Therefore, ManicTime data of the remaining four participants was used to complement qualitative explanations of their task switching behaviour, but it was not analysed quantitatively as previously planned to compare switching behaviour with and without the extension. Instead, the primary focus of data analysis was on the post-study interviews and participants' subjective experience of using the tools. The interviews were trancsribed verbatim, and analysed using thematic analysis.

6.3.4 Findings

Reflective information

Participants made to-do lists and schedules, but did not look back at them. P2 did look back at them, but only to move forward tasks he had not finished, and not to adjust expectations on his time management. They found it more useful to get information on current behaviour, as they did not know what to do with reflective information.

P3 would like to look back, and tried to with ManicTime, but felt it was effortful and did not have the time to reflect on it:

'I don't have time that I can set aside to work on things that will help me later (. . .). the only period where I would have the time to do these things, is when I get home. And when I get home, I'm at home, and I just can't be bothered doing more work stuff.' (P3)

Prospective information: to-do lists and schedules

Participants used prospective time information: they made written to-do lists but did not allocate time or deadlines to these. To-do lists were used to write down specific tasks without a time allocation. Schedules were used to allocate time blocks to higher-level tasks. P2 set aside several hours to work on 'finance-related' tasks, or email.

Type of interruptions

Digital versus non-digital interruptions Participants discussed that the information Focus provided would be useful, but that it currently did not capture the full data as it only looked at digital interruptions. Upon the first time of switching windows, the notification would say there was no interruption data yet:

'That's when I sort of thought: 'Oh, that's not really saying much, is it? Because it's not actually true. Because of course there were interruptions." (P2)

P1 and P2 said that they had just as much digital as non-digital interruptions, that these were equally disruptive, and felt that time data on digital switches only told part of the story. It would be useful to get information on both of these interruptions.

All participants found the timeline that ManicTime provided on time active and away from the computer useful.

P2 would use it to schedule in more breaks, whereas it made P3 realise his breaks were much longer than he thought:

'when I'm going to print something, and I end up talking with someone, and I come back, and it's like: you've been out for half an hour. And I was like: Jesus, time's flying! No wonder I work until 6 every day.' (P3)

Participants did not find a list of applications useful, as they spent the most time in the same applications. P2 and P3 only looked at it out of curiousity, but did not see the use of it for work.

Physical environment

Participants did not always work in the same location. For example, P3 worked in two different offices, and all participants worked from home occasionally. Different environments introduced different types of interruptions.

Gaps in data

X participants worked days from home, which left gaps in the ManicTime data.

Time feedback

Participants were particularly interested in the time it took them when they were not at their task. For P3, it helped him take shorter and fewer breaks:

'it made me realise how long I was spending, spending/wasting, doing other stuff. I think that affected me in the sense that I wanted to take fewer breaks. Because I didn't, well by breaks I mean, as I told you earlier, it's just going to do something, and then ending up chatting with someone in the passing.' (P3)

However, for other people it was not surprising:

'I don't really know whether that's helpful. To me, it doesn't kind of make me think: 'Oeh, I've been away too long'. I just think: OK, well I'm roughly aware that I've been away for an hour (. . .), I don't see how it links with being more productive. Unless I suppose, you're really easily distracted.' (P7)

Switching between windows

Participants did not find it useful to see how much time they spend in certain applications or windows. P2 was curious, but did not find it surprising as he spends most time in the information systems he uses for work, which are the main task windows. Most tasks revolved around a main task window, and participants found it easy to select one.

Distractions

6.3.5 Discussion

6.3.6 Limitations

6.3.7 Contributions

- Development of design recommendations for an expenses system.
- Demonstrate how an understanding of the used information sources and people's switching strategies between entering and looking up information can be used to adapt the design of the data entry interface.
- Demonstrate the applicability of design recommendations in the financial office settings in which the expenses task is currently done.
- Demonstrate that design features can influence people's strategies in entering expenses in a financial office setting.

Chapter 7

General Discussion

Chapter outline

This chapter summarises the findings from the seven studies. It discusses the contribution that the findings make to knowledge, the practical contribution, and discusses open questions and suggestions for future work.

7.1 Summary of findings

7.2 Contributions

7.3 Limitations and Future work

References

Altmann, E. M., Trafton, J. G., & Hambrick, D. Z. (2014). Momentary interruptions can derail the train of thought. *Journal of Experimental Psychology: General*, *143*(1), 215–226. doi: 10.1037/a0030986

Altmann, E. M., Trafton, J. G., & Hambrick, D. Z. (2017). Effects of Interruption Length on Procedural Errors. *Journal of Experimental Psychology: Applied*, *23*(2), 216–229. doi: 10.1037/xap0000117

Back, J., Brumby, D. P., & Cox, A. L. (2010). Locked-out : Investigating the Effectiveness of System Lockouts to Reduce Errors in Routine Tasks. In *Chi 2010.*

Back, J., Cox, A. L., & Brumby, D. P. (2012). Choosing to Interleave : Human Error and Information Access Cost. In *Proceedings of the 2012 acm annual conference on human factors in computing systems* (pp. 1651–1654).

Barchard, K. a., & Pace, L. a. (2011). Preventing human error: The impact of data entry methods on data accuracy and statistical results. *Computers in Human Behavior*, *27*(5), 1834–1839. Retrieved from http://dx.doi.org/10.1016/j.chb.2011.04.004 doi: 10.1016/j.chb.2011.04.004

Barchard, K. a., & Verenikina, Y. (2013). Improving data accuracy: Selecting the best data checking technique. *Computers in Human Behavior*, *29*(5), 1917–1922. Retrieved from http://dx.doi.org/10.1016/j.chb.2013.02.021 doi: 10.1016/j.chb.2013.02.021

Bardram, J. E., Bunde-Pedersen, J., & Soegaard, M. (2006). Support for activity-based computing in a personal computing operating system. In *Chi '06* (pp. 211–220). Retrieved from http://doi.acm.org/10.1145/1124772.1124805 doi: http://doi.acm.org/10.1145/1124772.1124805

Bi, X., & Balakrishnan, R. (2009). Comparing usage of a large high-resolution display to single or dual desktop displays for daily work. In *Proceedings of the sigchi conference on human factors in computing systems (chi '09)* (pp. 1005–1014). Retrieved from http://dl.acm.org/citation.cfm?doid=1518701.1518855{%}5Cnpapers2://publication/doi/10.1145/1518701.1518855 doi: 10.1145/1518701.1518855

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. Retrieved from http://eprints.uwe.ac.uk/11735/1/thematic{_}analysis{_}revised{_}-{_}final.doc doi: 10.1191/1478088706qp063oa

Brumby, D. P., Cox, A. L., Back, J., & Gould, S. J. (2013). Recovering from an interruption: Investigating speed-accuracy tradeoffs in task resumption strategy. *Journal of Experimental Psychology: Applied*, *19*(2), 95–107. Retrieved from http://discovery.ucl.ac.uk/1396456/

Byrne, M. D., & Bovair, S. (1995). A Working Memory Model of a Common Procedural Error. *Cognitive Science*, *21*(1), 31–61.

Chincotta, D., Underwood, G., Ghani, K. A., Papadopoulou, E., & Wresinski, M. (1999). Memory Span for Arabic Numerals and Digit Words: Evidence for a Limited-capacity, Visuo-spatial Storage System. *The Quarterly Journal of Experimental Psychology*, *52*(2), 325–352.

Cutrell, E., Czerwinski, M., & Horvitz, E. (2001). Notification, Disruption, and Memory: Effects of Messaging Interruptions on Memory and Performance. In *Proceedings of interact 2001* (pp. 263–269). New York, NY, USA: Springer.

Czerwinski, M., Horvitz, E., & Wilhite, S. (2004). A Diary Study of Task Switching and Interruptions. In *Proceedings of the sigchi conference on human factors in computing systems (chi '04)* (pp. 175–182). doi: 10.1145/985692.985715

Czerwinski, M., Smith, G., Regan, T., Meyers, B., Robertson, G., & Starkweather, G. (2003). Toward Characterizing the Productivity Benefits of Very Large Displays. In *Interact'03* (pp. 9–16). Retrieved from `http://research.microsoft.com/en-us/um/redmond/groups/cue/publications/interact2003-productivitylargedisplays.pdf`

De Luca, A., Langheinrich, M., & Hussmann, H. (2010). Towards understanding ATM security: a field study of real world ATM use. *Proc. SOUPS 2010*, 1–10. Retrieved from `http://portal.acm.org/citation.cfm?id=1837110.1837131` doi: 10.1145/1837110.1837131

Dearman, D., & Pierce, J. S. (2008). It's on my other computer!: computing with multiple devices. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 767–776). doi: 10.1145/1357054.1357177

Diemand-Yauman, C., Oppenheimer, D. M., & Vaughan, E. B. (2011, jan). Fortune favors the bold (and the Italicized): effects of disfluency on educational outcomes. *Cognition*, *118*(1), 111–5. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/21040910` doi: 10.1016/j.cognition.2010.09.012

Dumais, S., Cutrell, E., Cadiz, J., Jancke, G., Sarin, R., & Robbins, D. C. (2003). Stuff I've Seen: A System for Personal Information Retrieval and Re-Use. In *Sigir '03.* Toronto, Canada. Retrieved from `https://www.microsoft.com/en-us/research/wp-content/uploads/2003/01/siscore-sigir2003-final.pdf`

Evans, A. C., & Wobbrock, J. O. (2012). Taming Wild Behavior : The Input Observer for Obtaining Text Entry and Mouse Pointing Measures from Everyday Computer Use. In *Chi* (pp. 1947–1956). Austin, Texas, USA.

Furniss, D., & Blandford, A. (2006). Understanding Emergency Medical Dispatch in terms of Distributed Cognition: a case study. *Ergonomics*, *49*(12 & 13), 1174 – 1203.

Gong, Y., & Zhang, J. (2009). Effects of Information Displays for Hyperlipidemia. In *Human interface and the management of information. information and interaction* (pp. 503–512).

Gould, S. J., Cox, A. L., & Brumby, D. P. (2016). Diminished Control in Crowdsourcing: An Investigation of Crowdworker Multitasking Behavior. *ACM Transactions on Computer-Human Interaction*, *23*(3), 1–27. Retrieved from `message:{%}3C3DF7070D-92E4-400F-B288-C71BA25CC839@ucl.ac.uk{%}3Epapers3://publication/uuid/AC17DF2C-0715-49A0-A788-126336835C8E` doi: 10.1145/2928269

Gould, S. J., Cox, A. L., Brumby, D. P., & Wiseman, S. E. M. (2015). Home is Where the Lab is: A Comparison of Online and Lab Data From a Time-sensitive Study of Interruption. *Human Computation*, 45–67. Retrieved from `http://discovery.ucl.ac.uk/1470342/1/Gouldetal.-2015-HomeisWheretheLabisAComparisonofOnlinea.pdf` doi: 10.15346/hc.v2i1.4

Gray, W. D., & Fu, W.-T. (2004). Soft constraints in interactive behavior: The case of ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head. *Cognitive Science*, *28*(3), 359–382. doi: 10.1016/j.cogsci.2003.12.001

Gray, W. D., Sims, C. R., Fu, W.-T., & Schoelles, M. J. (2006, jul). The soft constraints hypothesis: a rational analysis approach to resource allocation for interactive behavior. *Psychological review*, *113*(3), 461–82. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16802878 doi: 10.1037/0033-295X.113.3.461

Green, R. a., Hripcsak, G., Salmasian, H., Lazar, E. J., Bostwick, S. B., Bakken, S. R., & Vawdrey, D. K. (2014, dec). Intercepting Wrong-Patient Orders in a Computerized Provider Order Entry System. *Annals of emergency medicine*. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/25534652 doi: 10.1016/j.annemergmed.2014.11.017

Grudin, J. (2001). Partitioning DigitalWorlds: Focal and Peripheral Awareness in Multiple Monitor Use. In *Proceedings of the sigchi conference on human factors in computing systems (chi '01)* (pp. 458–465). Retrieved from http://dl.acm.org/citation.cfm?id=365312 doi: 10.1145/365024.365312

Hanrahan, B. V., & Pérez-Qu, M. A. (2015). Lost in Email: Pulling Users Down a Path of Interaction. In *Chi'15* (pp. 3981–3984). Retrieved from http://delivery.acm.org/10.1145/2710000/2702351/p3981-hanrahan.pdf?ip=128.16.5.5{&}id=2702351{&}acc=ACTIVESERVICE{&}key=BF07A2EE685417C5.D93309013A15C57B.4D4702B0C3E38B35.4D4702B0C3E38B35{&}CFID=1018446498{&}CFTOKEN=44534115{&}{_}{_}acm{_}{_}=1513782721{_}94007b7bb008 doi: 10.1145/2702123.2702351

Healy, A. F., Kole, J. a., Buck-Gengle, C. J., & Bourne, L. E. (2004, sep). Effects of prolonged work on data entry speed and accuracy. *Journal of Experimental Psychology: Applied*, *10*(3), 188–99. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15462620 doi: 10.1037/1076-898X.10.3.188

Hollan, J. D., Hutchins, E. L., & Kirsh, D. (2000). Distributed cognition: toward a new foundation for human-computer interaction research. *ACM Trans. Comput.-Hum. Interact.*, *7*(2), 174–196. doi: http://doi.acm.org/10.1145/353485.353487

Holtzblatt, K., & Beyer, H. (2014). *Contextual Design: Evolved* (J. M. Carroll, Ed.). Morgan & Claypool Publishers.

Hutchins, E. L. (1995). How a Cockpit Remembers Its Speeds. *Cognitive Science*, *19*, 265–288. Retrieved from http://doi.wiley.com/10.1207/s15516709cog1903{_}1 doi: 10.1207/s15516709cog1903_1

Iqbal, S. T., & Horvitz, E. (2007). Disruption and recovery of computing tasks: field study, analysis, and directions. In *Chi '07* (pp. 677—-686). San Jose, California, USA. Retrieved from http://dl.acm.org/citation.cfm?id=1240730 doi: 10.1145/1240624.1240730

Janssen, C. P., & Gray, W. D. (2012, mar). When, what, and how much to reward in reinforcement learning-based models of cognition. *Cognitive science*, *36*(2), 333–58. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/22257174 doi: 10.1111/j.1551-6709.2011.01222.x

Katidioti, I., & Taatgen, N. A. (2013, sep). Choice in Multitasking: How Delays in the Primary Task Turn a Rational Into an Irrational Multitasker. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *56*(4), 728–736. Retrieved from http://hfs.sagepub.com/cgi/doi/10.1177/0018720813504216 doi: 10.1177/0018720813504216

Kawado, M., Hinotsu, S., Matsuyama, Y., Yamaguchi, T., Hashimoto, S., & Ohashi, Y. (2003). A comparison of error detection rates between the reading aloud method and the double data entry method. *Controlled Clinical Trials*, *24*, 560–569. doi: 10.1016/S0197-2456(03)00089-8

Koppel, R., Wetterneck, T., Telles, J. L., & Karsh, B.-T. (2008). Workarounds to Barcode Medication Administration Systems: Their Occurrences, Causes, and Threats to Patient Safety. *Journal of the American Medical Informatics Association*, *15*, 408–423. doi: 10.1197/jamia.M2616.Introduction

Li, S. Y. W., Blandford, A., Cairns, P., & Young, R. M. (2008). The effect of interruptions on postcompletion and other procedural errors: an account based on the activation-based goal memory model. *Journal of Experimental Psychology: Applied*, *14*(4), 314–328. doi: 10.1037/a0014397

Li, S. Y. W., Cox, A. L., Blandford, A., Cairns, P., Young, R. M., & Abeles, A. (2006). Further investigations into post-completion error: the effects of interruption position and duration. In *Proceeding cognitive science conference 2006* (pp. 471–476). Retrieved from http://discovery.ucl.ac.uk/13267/

Li, S. Y. W., Cox, A. L., Or, C., & Blandford, A. (n.d.). Dime and punishment : effect of error consequence on checking behavior. *Journal of Experimental Psychology: Applied*.

Lin, C.-J., & Wu, C. (2011). Factors affecting numerical typing performance of young adults in a hear-and-type task. *Ergonomics*, *54*(12), 1159–1174. doi: 10.1080/00140139.2011.622794

Lin, C.-J., & Wu, C. (2013, jan). Reactions, accuracy and response complexity of numerical typing on touch screens. *Ergonomics*, *56*(5), 818–31. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23597044 doi: 10.1080/00140139.2013.767384

Lin, C.-J., & Wu, C. (2014, sep). Predicting numerical data entry errors by classifying EEG signals with linear discriminant analysis. *Behaviour & Information Technology*(November 2014), 1–12. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/0144929X.2014.945962 doi: 10.1080/0144929X.2014.945962

MacKenzie, I. S., & Soukoreff, R. W. (2002). Text Entry for Mobile Computing: Models and Methods,Theory and Practice. *Human-Computer Interaction*, *17*(May 2015), 147–198. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/07370024.2002.9667313 doi: 10.1080/07370024.2002.9667313

Mangen, A., & Velay, J.-L. (2010). Digitizing literacy: reflections on the haptics of writing. In M. H. Zadeh (Ed.), *Advances in haptics* (pp. 385–402). InTech. Retrieved from http://www.intechopen.com/books/export/citation/BibTex/advances-in-haptics/digitizing-literacy-reflections-on-the-haptics-of-writing

Mark, G., Gonzalez, V. M., & Harris, J. (2005). No task left behind? In *Proceedings of the sigchi conference on human factors in computing systems (chi '05)* (pp. 321–330). Portland, Oregon, USA. Retrieved from http://dl.acm.org/citation.cfm?id=1054972.1055017 doi: 10.1145/1054972.1055017

Mark, G., Iqbal, S. T., Czerwinski, M., Johns, P., & Sano, A. (2016). Email duration, batching and self-interruption: Patterns of email use on productivity and stress. In *Chi'16* (pp. 1717–1728).

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. Retrieved from http://doi.apa.org/getdoi.cfm?doi=10.1037/a0014402 doi: 10.1037/a0014402

Morgan, P. L., & Patrick, J. (2012). Paying the price works: Increasing goal-state access cost improves problem solving and mitigates the effect of interruption. *The Quarterly Journal of Experimental Psychology*(November 2012), 1–19. doi: 10.1080/17470218.2012.702117

Morgan, P. L., Patrick, J., Waldron, S. M., King, S. L., & Patrick, T. (2009, dec). Improving memory after interruption: exploiting soft constraints and manipulating information access cost. *Journal of Experimental Psychology: Applied*, *15*(4), 291–306. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/20025416 doi: 10.1037/a0018008

Morgan, P. L., Waldron, S. M., King, S. L., & Patrick, J. (2007). Harder to Access, Better Performance? The Effects of Information Access Cost on Strategy and Performance. In M. Smith & G. Salvendy (Eds.), *Human interface, part i, hcii 2007, lncs 4557* (pp. 115–125). Berlin: Springer-Verlag.

Mueller, P. a., & Oppenheimer, D. M. (2014, apr). The Pen Is Mightier Than the Keyboard: Advantages of Longhand Over Laptop Note Taking. *Psychological science*, *25*(6), 1159–1168. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/24760141 doi: 10.1177/0956797614524581

Murphy, H. C., Chen, M.-M., & Cossutta, M. (2016). An investigation of multiple devices and information sources used in the hotel booking process. *Tourism Management*, *52*, 44–51. doi: 10.1016/j.tourman.2015.06.004

Nakata, T. (2014). Improving Human Reliability On Checking. In *Proceedings of international symposium on interaction design and human factors.*

Norman, D. A. (1993). Cognition in the Head and in the World : An Introduction to the Special Issue on Situated Action Cognition in the Head and in the World : An Introduction to the Special Issue on Situated Action. *Cognitive Science*, *17*, 1–6. doi: 10.1207/s15516709cog1701

O'Hara, K. P., & Payne, S. J. (1998). The effects of operator implementation cost on planfulness of problem solving and learning. *Cognitive psychology*, *35*(1), 34–70. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9520317 doi: 10.1006/cogp.1997.0676

Oladimeji, P., Thimbleby, H., & Cox, A. L. (2011). Number entry interfaces and their effects on error detection. In *Ifip international federation for information processing* (pp. 178–185). Springer-Verlag. Retrieved from http://discovery.ucl.ac.uk/1308862/

Oladimeji, P., Thimbleby, H., & Cox, A. L. (2013). A Performance Review of Number Entry Interfaces. In *Proceedings of 14th ifip tc13 conference on human-computer interaction (interact 2013), part i* (Vol. 8117, pp. 365–382).

Olsen, K. A. (2008). The $100,000 keying error. *Computer*, *41*. doi: 10.1109/MC.2008.135

Patrick, J., Morgan, P. L., Tiley, L., Smy, V., & Seeby, H. (2014). Designing the Interface to Encourage More Cognitive Processing. In *Hci* (Vol. 23, pp. 255–264).

Randall, D., & Rouncefield, M. (2014). Ethnography. In M. Soegaard & R. F. Dam (Eds.), *The encyclopedia of human-computer interaction* (2nd ed.). Aarhus, Denmark: The Interaction Design Foundation. Retrieved from https://www.interaction-design.org/encyclopedia/ethnography.html

Robertson, G., Czerwinski, M., Baudisch, P., Meyers, B., Robbins, D., Smith, G., & Tan, D. (2005). The large-display user experience. *IEEE Computer Graphics and Applications*, *25*(4), 44–51. doi: 10.1109/MCG.2005.88

Rule, A., Tabard, A., Boyd, K., & Hollan, J. D. (2015). Restoring the Context of Interrupted Work with Desktop Thumbnails. In *Proceedings of the 37th annual meeting of the cognitive science society* (pp. 2045–2050). Pasadena, USA.

Rule, A., & Youngstrom, K. (2013). That Reminds Me: Identifying Elements of Screen Recordings that Cue Contextual Memory. In *Cogs-230'13.* San Diego, CA, USA.

Rzeszotarski, J. M., Chi, E., Paritosh, P., & Dai, P. (2013). *Inserting Micro-Breaks into Crowdsourcing Workflows* (Tech. Rep.).

Salthouse, T. a. (1986). Perceptual, cognitive, and motoric aspects of transcription typing. *Psychological bulletin*, *99*(3), 303–319. doi: 10.1037/0033-2909.99.3.303

Salvucci, D. D., & Bogunovich, P. (2010). Multitasking and Monotasking : The Effects of Mental Workload on Deferred Task Interruptions. In *Chi 2010* (pp. 85–88).

Sandnes, F. E. (2013). An Error Tolerant Memory Aid for Reduced Cognitive Load in Number Copying Tasks. In *Universal access in human-computer interaction. user and context diversity* (pp. 614–623). Springer-Verlag.

Sellberg, C., & Susi, T. (2014). Technostress in the office: A distributed cognition perspective on human-technology interaction. *Cognition, Technology and Work*, *16*, 187–201. doi: 10.1007/s10111-013-0256-9

Smith, M. R., Lewis, R. L., Howes, A., Chu, A., & Green, C. (2008). More than 8,192 ways to skin a cat: Modeling behavior in multidimensional strategy spaces. *Proceedings of the 30th Annual Conference of Cognitive Science Society*, 1441–1446. Retrieved from http://www-personal.umich.edu/{~}rickl/pubs/smith-lewis-et-al-2008-cogsci.pdf

Sohn, T., Li, K. A., Griswold, W. G., & Hollan, J. D. (2008). A diary study of mobile information needs. In *Chi '08* (pp. 433–442). Retrieved from http://portal.acm.org/citation.cfm?doid=1357054.1357125 doi: 10.1145/1357054.1357125

Su, N. M., Brdiczka, O., & Begole, B. (2013). The Routineness of Routines: Measuring Rhythms of Media Interaction. *Human–Computer Interaction*, *28*(4), 287–334. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/07370024.2012.697026 doi: 10.1080/07370024.2012.697026

Thimbleby, H. (2011). Interactive Numbers : a Grand Challenge. In Blashki, K. (Ed.), *Proceedings of the iadis international conference on interfaces and human computer interaction* (pp. 1–9).

Trullemans, S., Sanctorum, A., & Signer, B. (2016). PimVis: Exploring and Re-finding Documents in Cross-Media Information Spaces. In *Proceedings of the international working conference on advanced visual interfaces - avi '16* (pp. 176–183). Bari, Italy. Retrieved from https://www.academia.edu/24295450/PimVis{_}Exploring{_}and{_}Re-finding{_}Documents{_}in{_}Cross-Media{_}Information{_}Spaces{%}5Cnhttp://dl.acm.org/citation.cfm?doid=2909132.2909261 doi: 10.1145/2909132.2909261

Tu, H., Oladimeji, P., Wiseman, S. E. M., Thimbleby, H., Niezen, G., & Cairns, P. (2014). Employing Number-based Graphical Representations to Enhance the Effects of Visual Check on Entry Error Detection Employing Number-based Graphical Representations to Enhance the Effects of Visual Check on Entry Error Detection. In *International symposium on interaction design and human factors (idhf 2014).* Kochi, Japan.

Vertanen, K., Memmi, H., Emge, J., Reyal, S., & Kristensson, P. O. (2015). VelociTap : Investigating Fast Mobile Text Entry using Sentence-Based Decoding of Touchscreen Keyboard Input. In *Chi 2015* (pp. 659–668).

Waldron, S. M., Patrick, J., & Duggan, G. B. (2011). The influence of goal-state access cost on planning during problem solving. *Quarterly Journal of Experimental Psychology*, *64*(3), 485–503.

Waldron, S. M., Patrick, J., Morgan, P. L., & King, S. L. (2007). Influencing cognitive strategy by manipulating information access. *Computer Journal*, *50*, 694–702. doi: 10.1093/comjnl/bxm064

Whittaker, S., Hollis, V., & Guydish, A. (2016). 'Don't Waste My Time': Use of Time Information Improves Focus. In *Chi'16* (pp. 1729–1738). San Jose, CA, USA.

Wiseman, S. E. M., Cox, A. L., Brumby, D. P., Gould, S. J., & Carroll, S. O. (2013). Using Checksums to Detect Number Entry Error. In *Proceedings of the 2013 conference on human factors in computing systems (chi-2013)* (pp. 2403–2406).

Zhang, J., & Wang, H. (2009, jan). An exploration of the relations between external representations and working memory. *PloS one*, *4*(8), e6513. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2714979{&}tool=pmcentrez{&}rendertype=abstract doi: 10.1371/journal.pone.0006513

Appendix A

Information sheet

The information sheet given to the participants in Study 1 is shown in Figure A.1.

You will be given a copy of this information sheet.

Title of Project: **Data entry in multitask settings**

This study has been approved by UCL Research Ethics Committee
[Project ID Number UCLIC/1415/001/Staff Brumby/Borghouts]

Name, Address and Contact Details of Investigators:

Dr. Duncan Brumby 8.23 Malet Place Engineering Building UCL Gower Street London WC1E 6BT United Kingdom brumby@cs.ucl.ac.uk	**Judith Borghouts** 8.18 Malet Place Engineering Building UCL Gower Street London WC1E 6BT United Kingdom judith.borghouts.14@ucl.ac.uk
+44 (0)20 7679 0689	+44 (0)20 7679 0693

We would like to invite you to participate in this research project. You should only participate if you want to; choosing not to take part will not disadvantage you in any way. Before you decide whether you want to take part, please read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or you would like more information.

Details of Study
This study aims to investigate how people perform number transcription tasks. You will be asked about your experience with transcribing numbers, and asked to perform a number transcription task, to show the researcher how you would normally perform this type of task. We will also audio record the interview for further analysis.
The interview will take approximately **30-45 minutes** and you will be reimbursed with a **£10** Amazon voucher as a token of appreciation for your participation.

It is up to you to decide whether or not to take part. If you choose not to participate, you won't incur any penalties or lose any benefits to which you might have been entitled. However, if you do decide to take part, you will be given this information sheet to keep and asked to sign a consent form. Even after agreeing to take part, you can still withdraw at any time and without giving a reason.

All data will be collected and stored in accordance with the Data Protection Act 1998.

Figure A.1: Information sheet

Appendix B

Consent form

The consent form used for Study 1 is shown in Figure B.1.

Title of Project: **Data entry in multitask settings**

This study has been approved by the UCL Research Ethics Committee [Project ID Number UCLIC/1415/001/Staff Brumby/Borghouts]

Participant's Statement

I ……………………………………………......................................

agree that I have

- read the information sheet and/or the project has been explained to me orally;
- had the opportunity to ask questions and discuss the study; and
- received satisfactory answers to all my questions or have been advised of an individual to contact for answers to pertinent questions about the research and my rights as a participant and whom to contact in the event of a research-related injury.

I understand that I am free to withdraw from the study without penalty if I so wish, and I consent to the processing of my personal information for the purposes of this study only and that it will not be used for any other purpose. I understand that such information will be treated as strictly confidential and handled in accordance with the provisions of the Data Protection Act 1998.

Signed: Date:

Investigator's Statement

I ……………………………………………………………………….

confirm that I have carefully explained the purpose of the study to the participant and outlined any reasonably foreseeable risks or benefits (where applicable).

Signed: Date:

Researcher notes, optional clauses (the participant should indicate yes/no to these and initial):

- I understand that my participation will be audio recorded, and I am aware of, and consent to, any use you intend to make of the recordings after the end of the project.
- I agree to be contacted in the future for invitations to participate in follow-up studies.
- I understand that the information I have submitted will be published as a report and I will be sent a copy. Confidentiality and anonymity will be maintained, and it will not be possible to identify me from any publications.
- I understand that I am being paid for my assistance in this research and that some of my personal details will be passed to UCL Finance for administration purposes.

Figure B.1: Consent form

Appendix C

Interview script

The interview script used for Study 1 is given below. This script only served to guide the interview, and does not contain all questions that were asked. Based on what the participant was saying, follow-up questions were asked.

Before the interview

- ensure participant is aware of purpose research
- explain what will happen
- informed consent
- ask for permission to audio record interview

Work

- Tell me something about your work (what do you do)
- How many hours per week (full-time/part-time)
- How long have you been working here (at this company)
- How long have you been doing this type of work

Number entry

- What activities do you do for work that involve transcribing numbers? e.g. filling in expenses, tax returns, setting up invoices
- How often do you do this (per day/week)?
- How many numbers is it roughly that you have to enter?
- How long do you usually take?
- What type of numbers? Usually same numbers, or can it be anything?
- Do you get to enter numbers that are different from your familiar format? e.g. 2,000 or 2.000; 9/15/14 instead of 15/9/14
- Do you deal with foreign currencies?

- Tell me something about how you enter these numbers
- When do you do these tasks? Immediately when you get them, or save them for later? Morning, afternoon?
- Does urgency/time pressure influence how you do the task (if so, how)
- Do you do them in-between other tasks or save a particular part of the day for it?
- Do you do all tasks all at once, or take rests in between? (if rests, what do you do? switch to another task, have a coffee, lunch, break, etc.)
- Do you feel that the way you enter it changes after a while? e.g. you get better at it so it kind of becomes automatic, or less mentally exhausting? Or is it the opposite, becomes more exhausting?
- Do you do other things as well during this task e.g. listening to music, attending to another task
- Do you sometimes have to briefly store numbers in memory, or calculate them from numbers you already have? If so, do you use external tools to offload memory?
- Where do you copy them from? Paper, digital files, combination?
- Do numbers get checked, to see if they're correct? Do you or anyone else check these numbers?
- Do you ever get entered numbers from someone else, that you then have to check if they are correct?
- What is your general experience with transcribing numbers? e.g. easy, boring, part of the job

Environment

- Do you always work in the same environment, or sometimes work in different places, such as at home, or when you're on the train, or working at a cafe? What about number entry tasks?
- Do you do your work on a desktop, laptop, tablet, anything else? Are some devices harder or easier?
- How is your desk organized?
- Do you organise it differently when doing number entry tasks?
- Do you have notifications on (e.g. e-mail, work-related instant messaging); if you do get new notification, do you attend to it straight away or finish task first?
- Do you get interrupted in other ways, for example when the phone is ringing, or when a colleague or your boss asks you something? How do you deal with these interruptions? What is your experience with these interruptions?
- Critical incident: Has there ever been an incident where a mistake in entering a number went undetected, and was discovered later on?

Demonstration

- Could you show me the software you use to transcribe numbers? What is your experience with this system, works well? (If negative, how do you deal with that? do you use any strategies to make it more optimal for yourself?)

- Do you feel confident entering the numbers?
- How do you place your windows?
- Could you show me how you perform a typical number transcription task (do it how you would normally); if you feel uncomfortable about sharing work data, you can enter any type of numbers, as long as it somewhat resembles data you would normally enter for work

After the interview

- Thank participant
- explain what will happen to their data
- do they have any more questions
- clarify when they will be compensated
- Ask if participant knows any further people who might be suitable and willing to participate