**\subsection{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.

%They may also change their strategies as they get more experienced with the task and know where to get the information from, or enter all information that is easy to access first.

The following questions will be addressed in this study:

\begin{enumerate}

\item

What is the information needed for an expenses task?

\item

Where is this information retrieved from?

\item

What are the strategies people use to look up information?

\end{enumerate}

%The models gave an insight in people's current task environment, and highlighted several issues. %which were grouped into three themes: lack of established coordination mechanisms, design of artefacts, and limitations in communication bandwidth. In order to identify patterns of information strategies the models, transcripts and notes, were reviewed and all findings that related to people's strategies were grouped in a separate category.

%

%\subsubsection{Expected findings}

%Based on findings from Study 1 and previous work on the influence of information access costs on task strategies \citep{Gray2006}, I expect the following findings:

%

%\begin{itemize}

%\item

%People will have to switch multiple times between different data sources as part of the same expenses task.

%\item

%The IAC of these data sources varies.

%\item

%If IAC of an information source is high, people will rely more on knowledge in the head: they will copy over more data in one go when IAC is high.

%\item

%Batching is a task planning strategy that gets employed by people with some experience in the task, in order to reduce switching between data entry tasks, and other tasks. People will save up data entry tasks to do them all in one sequence.

%

%Depending on people's experience and awareness of how costly it is to access information, people may plan their expenses task to reduce switching between entering and looking up information. They 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. An explanation for this is that they minimise the start-up costs of the entry task.

%\end{itemize}

**\subsubsection{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.

\textbf{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.

\textbf{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.

%Furthermore, collecting and organising information can be time-consuming (Bardram et al., 2006). This was also observed in the current study.

\textbf{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.

\textbf{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 \citep{Bi2009, Gurdin2001}. 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 \citep{Bardram2006}. 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.

\citet{Dearman2008}'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.

**\subsection{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:

\begin{enumerate}

\item

What is the information needed for an expenses task?

\item

Where is this information retrieved from?

\item

What are the strategies people use to look up information?

\end{enumerate}

%WHAT INFORMATION AND WHERE FROM

**\subsubsection{Information sources}**

%Paper and digital sources

%Availability of information

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.

%WHAT ARE THE STRATEGIES PEOPLE USE

**\subsubsection{Information strategies}**

%Collect physical sources in a batch, before starting a task, %Lookup digital information as they need it

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 \citet{Sohn2008}, 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.

%Use of multiple screens

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 \citep{Bi2009, Dearman2008, Grudin2001}. 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 \citep{Bardram2006}.

\citet{Dearman}'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 \citep{Grudin2001} where people deliberately used a second screen to look up information, even if it was faster to retrieve it from their primary screen.

%Use of local copies

**\subsubsection{Other issues}**

**\subsubsection{Contribution}**

\begin{itemize}

\item

Evidence that an expenses task is fragmented: people often have to go in and out of the expenses system to look up information.

\item

Evidence that the IAC of the required information sources for an expenses task varies.

\item

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.

\end{itemize}

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.

**\subsubsection{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.