**\section{Study 4: Copying data from multiple sources}**

**\subsection{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.

% Study 1 and 2 showed that for an expenses task, people have to collect the data to enter from multiple sources. Some of these sources are easy to access such as paper sheets on a desk, while others have a higher IAC, such as a computer document or window that takes time to open and view.

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 \citep{Cangiano2009, Murphy2016, Su2013}, 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.

%Study 3 has given an understanding of the effect of IAC on people's switching strategies when copying from one information source. As IAC increases, people make fewer visits to the source and instead enter what is in their head. The current study aims to investigate how differences in IAC affect people's strategies in switching between entering and looking up information from multiple sources, and whether different strategies affect performance.

%The office setting of Study 1 and 2 will be simulated in a laboratory environment. Participants will have to retrieve data from a number of sources, and enter this into a desktop computer using keyboard and mouse. The sources will be made to resemble the sources identified in Study 1 and 2 and will be on paper, on a second computer screen, or on the same screen as where the participant has to enter the data. The data participants have to enter will be similar to data that is entered for an expenses task: this includes names, financial numbers, and alphanumeric strings.

%Study 3 used eyetracking data and mouse movements to measure number and duration of visits to the target source. For the setup in this experiment, these measurements are not possible for visits to paper sources with the eyetracking equipment available. I will therefore video record and/or observe participants and code the timing and number of visits people make to the different information sources. To supplement these with quantitative measures, I will also measure keylogging data to get further insight whether and how long people interrupt entering data, presumably to retrieve data. In accordance with \citet{Gould2016}, intervals of more than 5s are considered to be an interruption.

%A within-subjects design will be used. The independent variables will be type of data (e.g. names, financial numbers, and alphanumeric strings), and IAC of the information sources. Dependent variables will be number of visits to sources, timing of visits, resumption lag, interkey intervals, typing speed, and error rate.

%The identified behaviour of how people manage looking up information is used to create a set of design recommendations for the current expenses system, which are evaluated in the next chapter.

The following questions will be addressed:

%\begin{itemize}

%\item[RQ1.] \textit{How does IAC affect the number of visits to look up information from multiple sources with different IACs?}

%\item[RQ2.] \textit{How does IAC affect the order of visits to look up information from multiple sources with different IACs?}

%\item[RQ3.] \textit{Do the number and order of visits affect data entry accuracy and speed?}

%\end{itemize]

The soft constraints hypothesis predicts that people choose and adapt their task strategies in order to minimise time \citep{Gray2006}. 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:

\begin{itemize}

\item[H1.]

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

\end{itemize}

In \citet{OHara1998}'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 \citep{Patrick2014}. 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:

\begin{itemize}

\item [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.

\end{itemize}

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 \citep{Monk2008}, and the more likely they are to make resumption errors \citep[e.g.][]{Brumby2013}. Based on this, the following hypothesis is made:

\begin{itemize}

\item[H3.]

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

\end{itemize}

**\subsection{Method}**

**\subsubsection{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 $\pounds$4 for their participation.

**\subsubsection{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 \ref{fig:ch34\_4-tasklayout}). 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 to use to reimburse the expense. They 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.

%The names of computer windows with an increased IAC were underlined, to make it easier to see which windows had an increased time cost to open.

\begin{figure}

**\includegraphics**[width=\textwidth]{images/ch34/ch34-4\_Tasksequence.pdf}

\caption{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).}**\label{fig:ch34\_4-tasklayout}**

\end{figure}

**\subsubsection{Materials}**

The numbers to be entered were made to resemble values that are ecologically relevant to an expenses 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 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).

%An experimental session consisted of 50 trials, divided into 5 blocks of 10 trials. Each trial consisted of four data entries, so in total a participant made 200 entries. 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. There were two practice trials before the experiment began. The numbers used in the practice trials were not used for the experimental trials. The data items always had the same length: the amounts consisted of two digits on the integer part and two digits on the fraction part (e.g. 11.94), and the account codes were six digits (e.g. 654273).

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.

**\subsubsection{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.

**\subsubsection{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.

**\subsubsection{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.

**\subsubsection{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, and get insight into the order in which people visited and entered data items. During the second iteration of grouping, for each trial the order of actions was considered and the trial was either grouped under a new strategy group for this order, or the trial was grouped under an existing strategy group.

%At the start of each experiment, some participants tried out different strategies. After several trials, most participants chose and stuck with the same strategy. Some participants used the same strategy for all trials. Therefore, the number of trials on which a certain strategy was chosen was more a categorical than continuous variable: people used it on zero or more than thirty trials.

**\subsection{Results}**

Table \ref{tbl:ch34\_4-means} 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.

\begin{table}

**\includegraphics**[width=\textwidth]{images/ch34/ch34-means.pdf}

\caption{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.}

**\label{tbl:ch34\_4-means}**

\end{table}

**\subsubsection{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 \ref{fig:ch34\_4-boxplots} 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\%.

\begin{figure}

**\includegraphics**[width=0.6\textwidth]{images/ch34/ch4\_4-boxplot.pdf}

\caption{Boxplot of interleaving rates in each condition.}

**\label{fig:ch34\_4-boxplots}**

\end{figure}

\begin{figure}

**\includegraphics**[width=\textwidth]{images/ch34/ch34-4\_linechart.pdf}

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

**\label{fig:ch34\_4-linechart}**

\end{figure}

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 visited 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 \ref{fig:ch34\_4-linechart}, which shows the distribution of interleaving rates for each condition. The lines all have peaks at the left and right end, indicating the interleaving rate was predominantly 0 or 100\% in each condition.

The consistency in strategy choice per participant is further illustrated in Figure \ref{fig:ch34\_4-plotpp}, 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, as the majority of plots have a flat line (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: at the start of the x axis, their lines go up and down between 0 (no interleaving) and 1 (interleaving) before becoming a straight line. Lastly, participants 29, 32 and 33 seemed to switch between the strategies throughout the experiment and did not stick with a particular strategy: their lines continue to go up and down between 0 and 1 along the entire x axis.

\begin{figure}[!htbp]

\centering

\begin{subfigure}[b]{0.5\textwidth}

**\includegraphics**[width=\textwidth]{images/ch34/ch34-4\_plotControl(1).png}

**\includegraphics**[width=\textwidth]{images/ch34/ch34-4\_plotControl(2).png}

\caption{Participants in the Control condition.}

\end{subfigure}

~ %add desired spacing between images, e. g. ~, \quad, \qquad, \hfill etc.

%(or a blank line to force the subfigure onto a new line)

\begin{subfigure}[b]{0.5\textwidth}

**\includegraphics**[width=\textwidth]{images/ch34/ch34-4\_plotHigh-Am(1).png}

**\includegraphics**[width=\textwidth]{images/ch34/ch34-4\_plotHigh-Am(2).png}

\caption{Participants in the High-Amount condition.}

\end{subfigure}

\begin{subfigure}[b]{0.5\textwidth}

**\includegraphics**[width=\textwidth]{images/ch34/ch34-4\_plotHigh-Acc(1).png}

**\includegraphics**[width=\textwidth]{images/ch34/ch34-4\_plotHigh-Acc(2).png}

\caption{Participants in the High-Account condition.}

\end{subfigure}

~ %add desired spacing between images, e. g. ~, \quad, \qquad, \hfill etc.

%(or a blank line to force the subfigure onto a new line)

\caption{A plot per participant across trials. The x axis shows the trial number, and the y axis indicates whether a participant interleaved on a trial: a value of 0 means they did not interleave, a value of 1 means they did interleave.}**\label{fig:ch34\_4-plotpp}**

\end{figure}

**\subsubsection{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 \ref{fig:ch34\_4-groupstr}. 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 \ref{tbl:ch34\_4-groupstr} shows the frequency with which these strategies were chosen per condition.

\begin{figure}[!ht]

\centering

**\includegraphics**[width=0.5\textwidth]{images/ch34/ch34\_4-groupstr.png}

\caption{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.}

**\label{fig:ch34\_4-groupstr}**

\end{figure}

\begin{table}[!ht]

\centering

\resizebox{\textwidth}{!}{

\begin{tabular}{l|l|l|l|l|ll}

\cline{2-5}

& \multicolumn{2}{l|}{Sequential} & \multicolumn{2}{l|}{Interleaving} & & \\ \hline

\multicolumn{1}{|l|}{Condition} & First row (a) & \begin{tabular}[c]{@{}l@{}}First \&\\ Second row (b) \end{tabular} & Amounts (c) & \begin{tabular}[c]{@{}l@{}}Amounts \& \\ Accounts (d) \end{tabular} & \multicolumn{1}{l|}{Other} & \multicolumn{1}{l|}{Total} \\ \hline

\multicolumn{1}{|l|}{High-Account} & 34\% {\footnotesize (48)} & 4\% {\footnotesize (6)} & 57\% {\footnotesize (80)} & 2\% {\footnotesize (3)} & \multicolumn{1}{l|}{3\% {\footnotesize (4)}} & \multicolumn{1}{l|}{100 {\footnotesize (141)}} \\ \hline

\multicolumn{1}{|l|}{High-Amount} & 20.99\% {\footnotesize (44)} & 16.57\% {\footnotesize (35)} & 49.72\% {\footnotesize (104)} & 9.39\% {\footnotesize (20)} & \multicolumn{1}{l|}{3.31\% {\footnotesize (7)}} & \multicolumn{1}{l|}{100 {\footnotesize (210)}} \\ \hline

\multicolumn{1}{|l|}{Control} & 11.2\% {\footnotesize (16)} & 21.6\% {\footnotesize (32)} & 54.4 {\footnotesize (81)} & 12\% {\footnotesize (18) } & \multicolumn{1}{l|}{0.8\% {\footnotesize (1)}} & \multicolumn{1}{l|}{100 {\footnotesize (148)}} \\ \hline

\multicolumn{1}{|l|}{Total} & 21.18\% {\footnotesize (190)} & 15.02\% {\footnotesize (73)} & 52.96\% {\footnotesize (265)} & 8.37\% {\footnotesize (41)} & \multicolumn{1}{l|}{2.46\% (12)} & \multicolumn{1}{l|}{100 {\footnotesize (499)}} \\ \hline

\end{tabular}

}

\caption{The most common grouping strategy was to chunk the items into three groups. The strategies are shown graphically in Figure \ref{fig:ch34\_4-groupstr}.}**\label{tbl:ch34\_4-groupstr}**

\end{table}

**\subsubsection{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\%, 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.

\citet{Wiseman2011} 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.

**\subsubsection{Qualitative findings}**

After the experiment had ended, participants were debriefed and the purpose of the study was explained. Some participants reflected on their strategies and gave additional explanations behind them. While these explanations are not the main focus of analysis and only serve to complement the quantitative measures, it helps understand people's motivation behind some of the measured strategies.

Participants mentioned they 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, five 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 items at a time. As a result, most participants ended up viewing and entering each account code one by one. This type of strategy is illustrated in Figure \ref{fig:ch34\_4-groupstr}.

Four 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 recognised, they would memorise this item as well as another item, before returning to the entry form. If they did not recognise the number, they would memorise one item. Furthermore, as data items had a fixed length, some participants started a trial by entering placeholders: they entered amounts of four digits and a decimal point, and account codes of six digits. They would then visit the information pages to check which of the digits of the items they needed to change.

**\subsection{Discussion}**

%Summary

The aim of this study was to investigate the effect of IAC on the order, number and durations of task interruptions to look up information from multiple information sources with varying IACs. While IAC did not influence the number and durations of task interruptions, it did influence the order in which people visited and entered 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.

%\item[RQ2.] \textit{How does IAC affect the order of visits to look up information from multiple sources with different IACs?}

%\item [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.

%Why would people continue to do it one by one if there is an added cost?

**\subsubsection{Order of data entries}**

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.

**\subsubsection{Chunking of data items}**

%\item[RQ1.] \textit{How does IAC affect the number of visits to look up information from multiple sources with different IACs?}

%\item[H1.]As IAC increases, people will try to group and memorise multiple items to minimise visits.

In contrast with prior work \citep{Gray2006}, an increase in IAC did not reduce the number of visits. However, in these prior studies there was no interaction involved to view information in the Low-IAC condition: information was displayed next to the data entry. In the current study, participants always had to move their mouse and click in order to view the information pages, which may have encouraged them to try and reduce visits and chunk items even in the Control condition. The added delay instead caused people which items to chunk, rather than whether they chunked or not.

%PERFORMANCE

%\subsubsection{Performance}

%\item[RQ3.] \textit{Do the number and order of visits affect data entry accuracy and speed?}

%The number of visits suggest participants made one visit per trial, and presumably this was the minimum number of visits necessary to enter the data items correctly. Even though a common strategy was to try and group both amounts, the average number of 4 visits per trial suggests participants either went back a second time and check they had entered both amounts correctly, or that they had to make more than one visit per account code.

%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 \citep{Morgan2014}, 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.

**\subsubsection{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.

In this study, both expenses were shown on the same page, and could be seen as part of the same task. Workers in Study 2 however not only dealt with multiple information pages, but also multiple data entry tasks. Based on the results of the current study, the hypothesis is made that a difference in IAC makes people more likely to interleave between data entry tasks, which is tested in Study 5.