Evaluation

**Participants**

Convenient samples of 8 participants were selected from amongst friends of the team members, of which 3 are females and 5 are males (?). All the participants are UBC students and are computer users with average level of computer knowledge. The first participant was intended to be a pilot subject, but since no significant changes were made in the study, the responses from the pilot study are included here. Four participants (P2, P3, P4 & P5) did the experiments in an observation room specially designed for human subject experiments (discussed below), and all other participants completed the experiments in general settings with one of the team members.

**Apparatus**

The experiments were conducted by using two connected personal computers. The software system of one of them was specially configured as the working environment for the participant; we call this computer the *participant’s computer*. The other computer was controlled by an experimenter to give simulated tasks and interruptions to the participants; we call this the *experimenter’s computer*.

The Participant’s Computer

The participant’s computer was running Windows 7 Professional as its operating system. To simulate the working environment of a regular computer user, we created a new clean user account on the computer. Preinstalled software includes Mozilla Thunderbird for receiving email messages, Windows Live Messenger for receiving instant messages, Mozilla Firefox and our TabFour prototype as the web browsers. In addition, we also installed screen capture software Fraps to record real screen motion of the participant for further analysis.

The Firefox web browser was ensured to have no bookmarks or plug-ins installed, which may influence the accuracy of our experiment. The contexts created in the TabFour browser were cleaned and the email messages in Thunderbird were also cleaned after each experiment. Fraps was turned on once the computer was handed over to the participant. We realized that Fraps may cause performance issues to the computer but this will not influence the results of the experiment. See detailed analysis of this in the pilot study section below.

The Experimenter’s Computer

The experimenter’s computers vary from the team member taking the experiment. While there is no special configuration regarding the experimenter’s computers, at least two pieces of software are installed. An email client was installed to send email messages to the participant’s computer and Windows Live Messenger was installed to send instant messages to the participant.

Due to the nature of the tasks, the experimenter pretended to be the participant’s friend whom is of the same gender, acting as either Mark or Jane. Separate email accounts were created specifically for this purpose, and the experimenter would send email messages and instant messages with the name of the virtual friend to the participant.

The Observation Room

We used the observation room on the 7th floor of the ICICS/CS building. The observation room is equipped with one-directional glass windows, allowing us to observe the actions of the subjects non-intrusively. The observation room has two video cameras fixed in two corners of the room. The experimenters outside the observation room can control the orientation of the camera, zoom in or zoom out remotely. The experimenters can also view the images of the two cameras on the same screen capture window. For the four participants completed their experiment in the observation room, we recorded tof the complete experiment to aid

**Procedure**

The experiment consists of using the two different browsers (Firefox, TabFour) to perform two different trip planning tasks (Orlando, London) that require the participants to search through multiple airline and tourist spot websites to locate cheapest fares and admission fees. One of the trips focused on planning a visit to London during spring break while the other was on a trip to Orlando during Christmas. We counterbalanced the order of the web browsers and the corresponding trips throughout our 8 participants, and we also ensured that different airline websites were used for each trip to avoid any possible learning effects. For each individual participant, the experiment was arranged as follows.

Tab4 Walkthrough and Experiment Introduction (5 minutes)

The participants first were given a brief walkthrough of the TabFour browser. One experimenter demonstrated the most common features of the browser but using the features or not was completely up to the participant. The experimenter would briefly introduce the procedure of the experiment, but the details of the tasks the participant was supposed to complete were briefly glanced over.

Task Sets (15 minutes per task)

In a task set using either of the two browsers, the participants first were given an email from the virtual friend, explaining that they are planning a trip and expecting the participants to look up the lowest airfares in a given time window. This is a relatively complex task involving opening multiple airline websites and comparing prices of different dates that is intended to keep the participant occupied throughout the time window. At about the 5th minute from the beginning of the airfare task, the experimenter sent out an instant message through the virtual friend to the participant, asking the participant to lookup the author names of some papers. The experimenter also stated the task was urgent and should be done immediately, to force the participant to interrupt current task and switch to a completely irrelevant context. Right after the participant sent back the results of the second task, the experimenter will then sent another IM message, requesting the participant to look up the prices for some places of interest of their trip destination. The participants were requested to send back the results of both the airfare task and the ticket task in one email at the end. The same procedure was repeated in the second task with the alternate browser and trip destination.

Evaluation Surveys (5 minutes each)

At the end of each task, the participants were asked to fill out an evaluation survey about how they feel when completing the tasks in the web browser just used. We are mostly interested in the workload required by the design of the primary trip planning tasks and the interruptions, and how well the browser they used supported what they were trying to achieve. After the completion of the two tasks, the participant was also asked to fill up a questionnaire that compares the two browsers. The details of these survey and their results have been discussed in the extended abstract write-up and the results section of the appendix.

**Pilot Studies**

We conducted two pilot studies. The first one is an informal experiment done by one of the experimenters, which we did not include this in the final result analysis. The second pilot study was done on a real participant in a formal setting and was included in the final results analysis, but the experiment was not carried out in the observation room.

First Pilot Study

Originally, we plan to monitor the screen of the subject through remote desktop software. By doing this, we believe only posed very little influence to the subject’s computer. We used TightVNC which is a free remote desktop monitor application. Then we planned to capture the screen of the subjects at the experimenter’s computer. However from the first pilot study, we discovered using remote desktop monitoring method has two major defects. First, the network bandwidth was mainly consumed by the remote desktop application and made web browsing noticeably slow. Second, remote desktop monitoring had a very low refresh rate. The quality of the video is not acceptable. Therefore we decided to record the screen on the participant’s computer and we found that by doing so, the performance impact was acceptable.

Another change we made from the first pilot study was the order of the interruptions. At first, we planned to give the participants the travel planning first, then the ticket task followed by the paper task at last. However, we found the paper task’s interruption effect was not strong enough, as the user was almost finished with the first two tasks by the time the paper task came at the 10th minute mark. Although the paper task was marked “urgent”, the users were likely to do it after completed the first two tasks, rendering the interruption meaningless. Therefore we exchanged the paper task with the ticket task and found the two interruptions worked well because although when the ticket task came, it was also approaching the end of the airfare search. As they were more closely related, the users were likely to do them simultaneously.

Another problem we discovered was some websites could not be rendered correctly in TabFour, because some websites were not compatible with WebKit, which is the render engine used in our browser. We eliminated these websites from the tasks and replaced them with alternative websites.

Second Pilot Study

The second pilot study was done on our first participant, under the revised experiment settings and procedures after the first pilot study. After the second pilot study, we made another minor correction to our experiment. Originally for the TabFour browser, only one task of the three tasks used the “online context sharing” feature. For the other two tasks, the user still had to manually copy and paste the link to the browser. Even worse, when the user was selecting the links, sometimes he clicked on a link accidentally, and the operating system opened the default browser (Firefox) for the link. To fix this and to better evaluate user reactions of the sharing context feature. We decided to change all three tasks to use shared contexts by sending the participants only the file name of the contexts to be loaded, hiding all explicit links.

**Study design**

**Dependent measures**

**Hypotheses**