Evaluation

**Participants**

The participants of the experiment are chosen randomly from the friends of the team members, based on their availability. We had 8 participants in total, 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 part of the pilot study. Participants 2, 3, 4 and 5 did the experiments in an observation room (discussed below) and all other participants completed the experiments in general settings with one of the team members.

**Apparatus**

The experiments were done on 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*. Four participants did the experiment in an observation room specially designed for human subject experiments.

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 Firefox, Mozilla Thunderbird for receiving email messages, Windows Live Messenger for receiving instant messages and our TabFour browser. 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.

The experimenter pretended to be a friend of the participant with the same gender, namely Mark or Jane. 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 glasses allowing us to observe the actions of the subjects but they will not be influenced by the movements of the experimenters outside the observation room. 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. The video stream can be recorded for further analysis. For the four participants completed their experiment in the observation room, we recorded the video of the participants as long as screen capture for analyzing user behaviors.

**Procedure**

We had two different set of tasks. The first one focused on planning a trip to London during spring break and the second task was focused on a trip to Orlando, Florida during Christmas. Each set of tasks was supposed to be completed in a different browser, either Firefox or TabFour. But the combinations of task sets and browsers and their orders vary among participants. For each individual participant, the experiment was arranged as follows.

TabFour 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 feature or not was completely up to the participant. Then the experimenter would briefly introduce the procedure of the experiment, but he would not touch on the details of the tasks the participant was supposed to complete.

Task Set 1 (15 minutes)

The participant will then begin his/her first set of tasks. The first set of tasks was on either Firefox or TabFour. We arranged our participants to have different orders of browsers, to counteract the influence of always using one browser before another.

In one task set, the participants first were given an email, from the virtual friend, telling him/her 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.

Then at about the 5th minute from the beginning of the airfare task, one experimenter sent out an instant message through the virtual friend to the participant, saying he/she was doing some research stuff and wanted 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.

Evaluation Survey 1 (5 minutes)

The participants then were asked to fill out an evaluation survey about how they feel when completing the tasks in the web browser just used.

Task Set 2 (15 minutes)

Then the participants were asked to do another set of tasks in a different browser. The content of the tasks were different but the tasks are of the same type with task set 1, i.e., both were airfare search followed by authors look up and then tickets look up. The procedures were the same as in Task Set 1.

Evaluation Survey 2 (10 minutes)

The participants then were asked to fill out an evaluation survey like the first evaluation survey about how they feel when completing the tasks in the web browser just used. Then they were asked to fill out another survey comparing their preference between the browsers.

**Pilot Studies**

We did two pilot studies. The first one is an informal experiment did on ourselves. 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. In 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, first 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, and the paper task at last. However, we found the paper task could not have the effect as an interruption. When the paper task came, the user was most probably wrapping up the first two tasks. So although the paper task was marked “urgent”, the users were likely to do it after completed the first two tasks and thus rendered the interruption meaningless. Therefore we exchange 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, they were closely related, so 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 then eliminate 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 setting and procedure after the first pilot study. After the second pilot study, we did one major change 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 the link accidentally and the operating system then opened default browser for the link, which is not our TabFour browser. 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, hiding all explicit links.

**Study design**

**Dependent measures**

**Hypotheses**