Assignment 5: Final evaluation

Study Description

Pilot Study and Improvements

Using the evaluation plan created in assignment 4, a pilot study was conducted with a peer during the lab. Tasks outlined in the plan were performed, and each step went smoothly in terms of assessing the evaluation plan. During discussion of our evaluation plan with our peer, they commented that it was detailed and coherent. Therefore, the pilot study proceeded satisfactorily.

During the pilot study, while there was little feedback on polishing our evaluation plan, we did receive actual feedback on the application itself:

- After choosing an answer on the questionnaire, the user is required to scroll down to confirm their choice. As a result, the user was perplexed on how to proceed. As seen in Figure 1 below, there is no visual indication of a scrolling function.

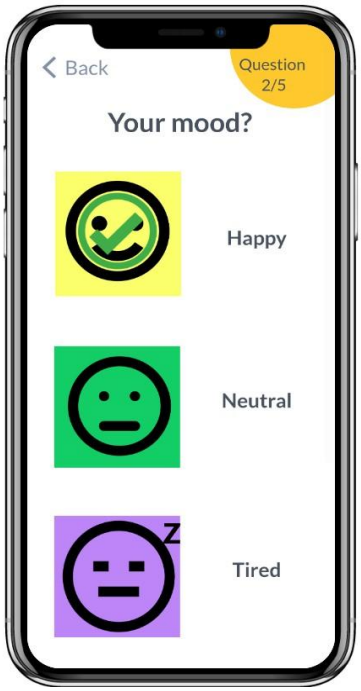

Figure 1: Scrolling feature on one of the questions is not visible.

- In the last step, the app takes the user to a third-party app to finish their order. Currently, there is no way to return to QuickDrink after the order has been

processed, and researchers had to step in to correct the path during the study. In Figure 2 below, this final step for ordering is shown.

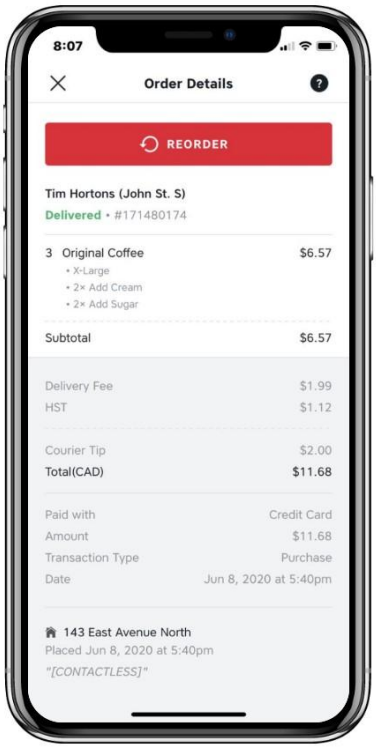

Figure 2: Third party ordering app with no way of returning to QuickDrink.

Participants Summary

For our study 5 participants were recruited to test our medium fidelity prototype. These participants are all in their 20s to 30s. They all have experience ordering drinks in-person and online. They also all have experience with phone applications. Only one participant consented to recording and photographing during the evaluation. Participants 2, 3, and 4 have a job while participants 1 and 5 are students at UVIC. Participants 2 and 4 usually get the same drink when ordering one.

First participant (P1) is a fourth-year computer science student in their early twenties. P1 is someone who faces choice overload by becoming overwhelmed when ordering food, including picking between drinks at a beverage shop.

Second Participant (P2) is a college student working a part-time job. P2 enjoys various ordering drinks, but has a tendency to stick to their favorites.

Third participant (P3) is a software developer. P3 works from his home office and travels to their work office a few times a week. P3 regularly places orders on apps.

Fourth participant (P4) enjoys beverages all day. P4 is a millennial in their late 20s who works in the gaming sector, and requires continuous refilling to stay energized.

Fifth Participant (P5) sometimes enjoys beverages more than 4 times a day. They prefer to drink in-store such as StarBucks.

Method Used

In this study, a moderated, co-discovery learning method was used, with researchers acting as the "coach." The researcher only acts as a guide for the participant and they do not interact with the prototype. Afterwards, the user is asked to participate in an unstructured interview with the researcher.

A sample set up of what the study looked like is seen in figure 3 below:

Figure 3: Research set up with the medium-fidelity prototype on-screen.

With the prototype, the participants were asked to complete three tasks:

- *Task 1:* Log-in to the application for the first time and receive a drink recommendation by going through the questionnaire.
- *Task 2:* Quickly receive a beverage recommendation by clicking the “Surprise me” button.

- *Task 3*: Leave a comment on previously ordered drinks by viewing a history of past recommendations.

Within each task are specific step-by-step instructions on how the tasks are performed within the prototype. Only researchers have access to these steps (can be found in assignment 4—Medium-Fidelity Prototyping and Evaluation Plan) which serve as the participant's guide.

After going through the tasks, the researcher asked the participant questions in an unstructured interview regarding their experience of using the app.

This approach allows the researcher to gather feedback in real time as the user interacts with the application. The researcher will be able to answer any questions the participants may have during the process, but more importantly, the researcher will be able to observe why they asked that question to begin with. This allows the team to gain real-time insight on how to further develop the application. Further, this allows the researcher to ask questions specific to the user and about events that occurred specifically during each evaluation.

Data Collected

In this study, the primary information collected are observations on user behavior and participant comments on the usability of the interface. These comments that criticize or praise the design decisions made during development are currently the most valuable. We also want to see that the questions posed in the app's questionnaire were relevant for the drink recommendation process. Through the unstructured interview, users shared their thoughts on any functionality that the app might be missing or lacking in substance.

Since emphasis on participants' feelings was placed for the evaluation, and by the nature of the information we wish to collect, this study produced mainly qualitative data. To further explain, our application deals with people's feelings regarding choice overload, so it is most important for us to collect qualitative data about our interface. Given instructions, participants engaged with the prototype. Meanwhile, data was collected by writing down observations. Further, more data was gathered by asking questions and elaborating on comments made during the unstructured interview portion. When participants were confused, the researcher provided assistance while maintaining minimum interruption to obtain more accurate data.

Our research naturally doesn't require much quantitative data. However, since QuickDrink is an app that users expect to be quick, the time taken for participants to

finish the questionnaire could have been measured. This is further discussed under "Limitations and Reflections."

Results

In this section, we present interesting observations and comments made by participants during the study. Our results are partitioned by tasks performed, and under each are insights we found intriguing.

Task 1

Task 1 demonstrates users using the app for the first time, answering the questions, getting a recommended drink, and finishing with ordering.

- The sign-in page confused one of the participants because they did not know whether to click the "OK" or the "Sign-in" button.
- Two users did not know how to proceed past the allergies page, as they had no allergies.
- For two participants, it was not obvious that they were able to scroll through answers on some questions.
- A participant pointed out that, when choosing a store to order from, stores should not pop up until one was selected on the presented map.
- Three participants pointed out the lack of direction after ordering the drink from a third-party app. (i.e. there was nowhere to go)

Task 2

Task 2 shows the functionality of the 'Surprise Me' button, where users receive a random recommendation without answering the questions.

- Two participants pointed out that the label "Surprise Me" intrigued them, but the phrase "Surprise Me" was not intuitive.

Overall, the participants completed this task quickly.

Task 3

Task 3 focuses on side features of the app such as the history page, and leaving ratings and comments on previously ordered drinks.

- One participant pointed out that the history button is not obvious.

- On the history page, one participant was not sure where to click to access previous drinks.
- A participant mentioned that using the word “reorder” to access a previous drink is bad phrasing.
- Two participants pointed out that the placement for rating a drink when leaving a comment is not optimal. Also, they were unsure of the next step after leaving a comment, since there is no obvious enter or confirm button.
- A participant also identified that there were no ratings associated with other comments.

Generally, all participants included good feedback throughout the process. Most of them related to some intuitive design and comfortability of the application. A common compliment from the participants was the well thought out process of the application. It shows that the overall flow of the prototype is well-received.

Discussion

With the results found above, and after a coding exercise, we formed three main themes through qualitative data analysis:

- Lack of information
- Clunky UI
- Ambiguous phrasing

Lack of Information

Participants drew attention to the application's lack of information. Participants found several buttons un-tactile and had difficulty scrolling after answering a question. Users also struggled with figuring out the next step after ordering a drink from the third-party app. A tutorial guide for first-time users would solve this problem. As a result, users should be well informed about the purpose of the buttons, when to scroll through a page as a result, and the direction to take after ordering.

Clunky UI

Participants also struggled with the clunky UI. The user interface at some points of the prototype was puzzling. For instance, the sign-in page featured both an "ok" and a "sign-in with Google" option causing confusion on how to proceed. Participants also commented on the allergens page being unsure of how to proceed if they don't have any allergies. An interface redesign on appropriate pages should resolve these issues.

Ambiguous Phrasing

Participants were also confused by some of the wording and phrasing used, such as "Surprise me" and "Re-order." They could not articulate the direct meaning of the words. "Surprise me" is intriguing but not intuitive, and a participant was concerned that pressing "Re-order" would automatically order the drink. Altering the button's text or including the functions of these buttons in the tutorial are two possible solutions.

Study Limitations and Reflections

Limitations

The study had three main limitations.

The first is a small number of participants. While our participants did provide valuable insight into our product and implementing HCI properly, we found that the number of participants limited the range of feedback received.

Second, there was a lack of diversity in our study. All our participants vary within their 20s to 30s, which is a very small age range for a study group. The participants are all also residents of the same city and have experience with ordering drinks and using an app.

The third limitation is that we had minimal ways of incorporating meaningful quantitative data. All the data we collected was qualitative data, such as their opinions, answers to questions, and observation notes. While this was the data we were searching for, having quantitative data gives us something more concrete, such as the time taken for a user to complete the questionnaire.

Reflections

Overall, the study provided crucial feedback and revealed which parts of the application needed more work. The most valuable insight being not to assume what our end-users already know. Throughout the study, users pointed out ambiguities within the app we thought were intuitive. We were also able to collect the qualitative data we needed, accomplishing the study's purpose.

In terms of what did not work well, the participant comments were not consistent. The participants either gave feedback for a specific task or more generally on the app. This can be explained as a general misunderstanding within the team on how to conduct the

study and what questions to ask. We also had to interrupt participants during the task process to ask them for elaboration on certain comments. As a result, most of the interview occurred during the tasks themselves rather than being a post-interview. This was done to receive real-time feedback rather than the elaboration being an afterthought.

The main improvement for this study is to increase the number and diversity of the participants. This would allow a more accurate portrayal of a larger variety of users. This also gives more feedback on the prototype. Another improvement would be to use a semi-structured interview format instead of unstructured. This ensures more consistency across the studies while still providing improvisation for the researcher. For further consistency and to avoid interruptions, any requests for elaborations or questions during the tasks should be saved for the end. The study could also be opened up to include ways to incorporate quantitative data like the amount of time participants took to receive and order a recommendation.