CS 6750 Human-Computer Interaction
Assignment M4

Waihong Leong
wleong8@gatech.edu

***Abstract*—With the onset of COVID-19, online grocery shopping with curbside pickup has gained wide adoption for consumers. Walmart is the leader in this space, and one of the first to offer a safe and convenient way of shopping grocery online through their mobile ap. The app interface has gone through multiple upgrades, but users today still find that it is overly complicated to perform many of the basic tasks. One such task is re-ordering frequent items from a shopping list or prior purchases. This project will focus on the unmet needs for this task on Walmart app. The goal is to re-design a more user-friendly interface for routine grocery shoppers.**

1 QUALITATIVE EVALUATION

Interviews will be conducted to evaluate Prototype#2 Easy Auto Reorder Interface (paper prototype), shown in Appendix A.

1.1 Evaluation Plan

The interviews will be **synchronous**, evaluating a single prototype interface. My plan is to recruit a total of 4 participants from family members and neighbors who are experienced online grocery shoppers. To avoid *groupthink*, interviews are conducted individually. Each session is scheduled for 30 minutes via face-to-face meeting and will take place either at my home or at the neighbor's home. The sessions will not be recorded. Instead, all data is captured via **note-taking**. This is to ensure participants feel totally comfortable to share their thoughts and feedback.

1.2 Evaluation Content

Prior to the interview, I will provide a quick introduction and some background context to the participants about the research as discussed in the abstract. I will briefly describe the purpose of the prototype and the associated task, which is to re-order grocery items via the Walmart mobile app. I will then present the paper prototype and allow the participants to look at it for about 3 minutes. The prototype is kept visible during the entire discussion for easy reference.

The interviews will be **semi-structured** instead of 100% scripted. This allows the flexibility to prop for more insights that are helpful to better understand the participant thought process and their needs. This is also the reason why I choose to conduct this evaluation with interviews instead of a survey.

Here is a list of open ended questions to ask the participants:

- **Q1:** What do you think about this interface? How useful or relevant is this to you?
- **Q2**: How easy or difficult for you to understand the functionality of the interface? Did you find anything confusing?
- **Q3:** How easy or difficult for you to use this interface? Would you enjoy using it?
- **Q4:** What do you like and dislike about this interface?
- **Q5:** What can be improved? Are there any features you would like to see that are not here?
- **Q6:** Would you use this interface? Why or why not?

1.3 Assessment on Design Requirements

The goal of the interviews is to evaluate whether the prototype has met the requirements that were identified in the needfinding phase (see Appendix D: Design Requirements). Q1 and Q5 addresses *functionality* of the prototype. We want to gauge how much the prototype is valued and whether it serves the needs for the users. We also want to know if any important features are missing that prevent the user from carrying out the grocery ordering task successfully. Q2 is meant to evaluate *learnability*. The interface should be intuitive enough to operate without any additional instructions needed. Any confusion must be resolved to

bridge the gulf of evaluation. Q3 and Q4 concern about *usability*. This also relates to user satisfaction. Difficulty of using the interface may indicate high *cognitive load* and require further improvement to simplify the process.

2 EMPIRICAL EVALUATION

Empirical evaluation will be conducted on Prototype#1 Voice Assistance for Re-ordering Grocery, shown in Appendix B. The goal here is to evaluate the efficiency of re-ordering grocery items via a voice agent vs. the current interface on Walmart app to decide which is better.

1.1 Control and Conditions

For this evaluation, I am interested in *efficiency*. I am testing the *length of time* required to complete the task of re-ordering grocery items using the voice interface prototype vs. current Walmart app interface. I want to know if the voice interface can help the user reduce time spent on this task. Here, we have two **conditions** which are called the *treatments*. The first treatment is the *voice interface prototype*. The second treatment is the *current Walmart app*, which also serves as the **control**.

1.2 Null and Alternative Hypothesis

Null hypothesis assumes that there is no statistical difference between the two treatments. For this experiment, my *null hypothesis* is that there is no difference in task completion time for re-ordering grocery items between the voice interface prototype vs. current Walmart app. The *alternative hypothesis* is that there is a difference in task completion time for re-ordering grocery items between the voice interface prototype vs. current Walmart app.

In general, we want to prove that one of the interfaces will take less time to complete the task. Therefore, one interface has better efficiency than the other if the alternative hypothesis is true.

1.3 Experimental Method

I will run a **within-subjects** experiment where each participant in the experiment will be given both treatments of the design interfaces. I choose *within-subjects* for two reasons. First, this helps to maximize the data we can collect with a smaller pool of participants. Second. This method also allows me to see how each individual participant is affected, rather than just the group as a whole. It is beneficial to observe how the same participant might perform between the two treatments to gain subtle insight at individual level.

To control biases, the order of treatments given to the participants will be **randomized**. Each participant is assigned an interface randomly to start with: current Walmart app or the prototype. For each treatment, the participant will complete the same task which has been standardized. The task involves ordering items from a predefined favorite grocery list with 3 items: 1 gallon of milk, 1 dozen of eggs and 1 loaf of bread. Measurement taken is the *total time* to complete the task with each interface. The type of data collected is ratio.

For data analysis, I will use **paired t-test** because we have two sets of *continues data*. This test will compare the means of two paired sets of data, since each participant is given both treatments.

1.4 Lurking Variables

Below, I will discuss some potential *lurking variables* that are not measured in the study but could affect the overall results of the experiment. To deal with these problems, it is important that we randomize the assignments and ensure that we have big enough samples size to draw statistical conclusions.

- **Increased familiarity** of the task. This is also known as *practice effect*. Since the participant is repeating the same task on both treatments, this could lead to slight advantage for the second treatment because the participant already knows what to do.
- **Fatigue** may impact the speed and reaction time of the participant to execute a task. We need to watch for fatigue and provide resting time in between treatments.

- Each participant might come with **different levels of experience**. This could greatly impact how quickly they can complete the task with the interface. To mitigate this, we need to ensure the recruitment is not over-weight in one type of users.
- Our **ability to administer** the experiment improves over time. This may lead to generally a smoother experience for the participants who join later with better performance.

3 PREDICTIVE EVALUATION

For predictive evaluation, I will construct a **GOMS model** to evaluate Prototype#3 Buy Again Interface (card prototype), shown in Appendix C.

The design of this interface is intended for **expert users** who are regular online grocery shoppers. We want to improve efficiency of a *known task*, which is to reorder grocery items that are frequently purchased. So, a model-based evaluation is a good fit for this purpose. We will use this model to analyze and compare the steps required to determine if the new interface actually helps to simplify the process and reduce execution times for the users.

Here is how the GOMS model will be constructed for evaluation:

- Focus on a **specific task**: re-ordering the same grocery items from a prior purchase.
- This task can be accomplished in two ways: the prototype interface vs. the current Walmart app interface. The **selection methods** will show both options.
- The **user goal** is to re-order grocery items with the least amount of time, in which the user already know how to perform.
- These **operators** are available to the users
    - Log into the app
    - Click the *'buy again'* button
    - Select *item(s)* to re-order
    - Click the *'add'* button to build basket
    - Click *'check out'* button
    - Select *time and date* to schedule pickup
    - Click *'place order'* button to complete checkout process

- Each operator is assigned an **execution time** to indicate how long it takes to perform the action.

Once the GOMS model is constructed, we can then compare and analyze each selection method to see which option takes the least amount of time to complete the task. We can also compare the number of operators which reflect the complexity of actions for each procedure. The less operators, the easier for the users to execute, and it is most likely to require less time to perform.

.

4 PREPARING TO EXECUTE

For the next assignment, I have selected the following two evaluations to perform.

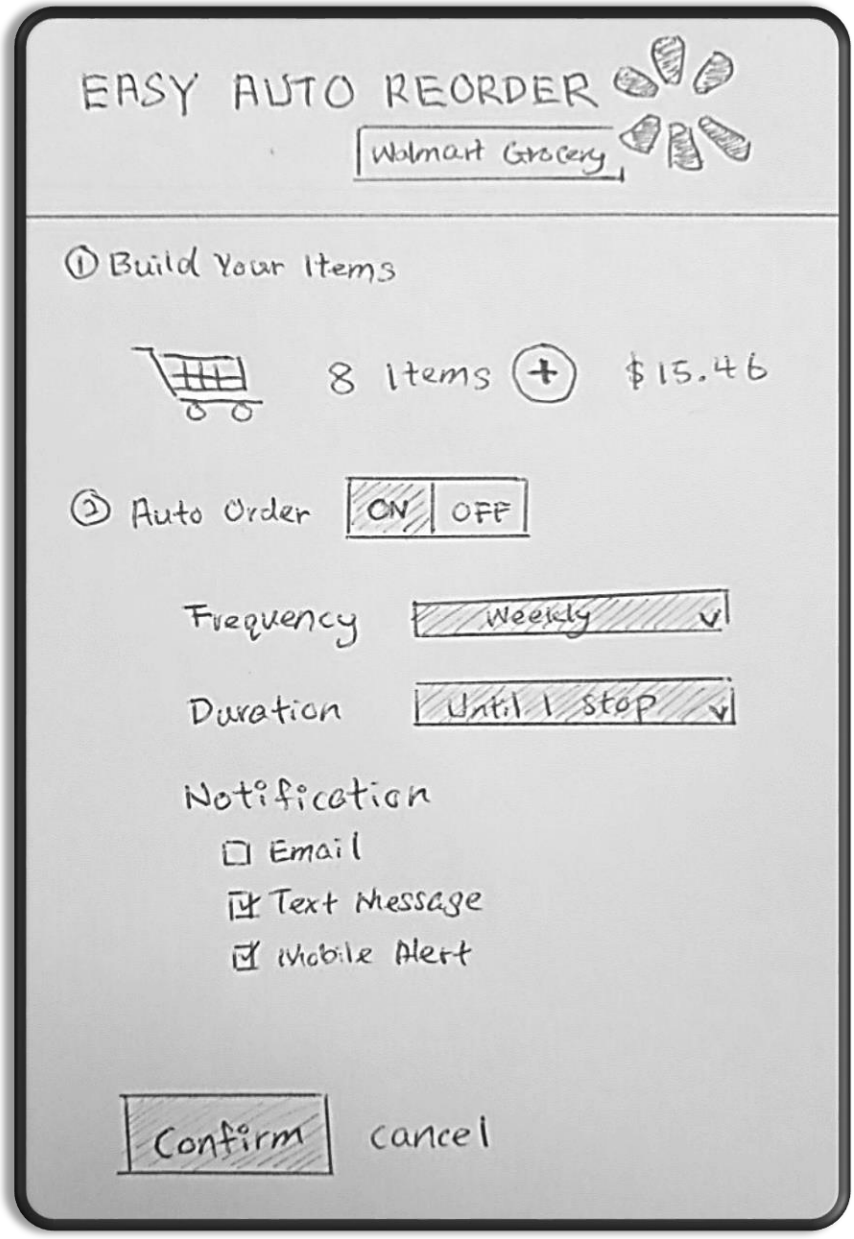
1) Qualitative evaluation with Prototype#2: Easy Auto Reorder Interface
2) Predictive evaluation with Prototype#3: Buy Again Interface

Here are the reasons why I selected these two for evaluation, but not the empirical approach:

- At the current stage, I am still early in the design process. I want to focus on **formative** evaluation. The primary purpose is to gather feedback to help redesign and improve the prototypes. Because of this reason, qualitative inputs are more helpful than quantitative data from a formal, empirical evaluation.
- Both qualitative and predictive evaluations are more suited for low fidelity prototypes which are what I have.
- All three of my prototypes are non-functional. They are not ready to be subjected for actual use in an empirical experiment to measure performance data.
- Lastly, empirical evaluation requires a control study with larger sample size of participants to derive statistical conclusions. The cost is too high, and it outweighs the benefits we can gain from this study at this stage.

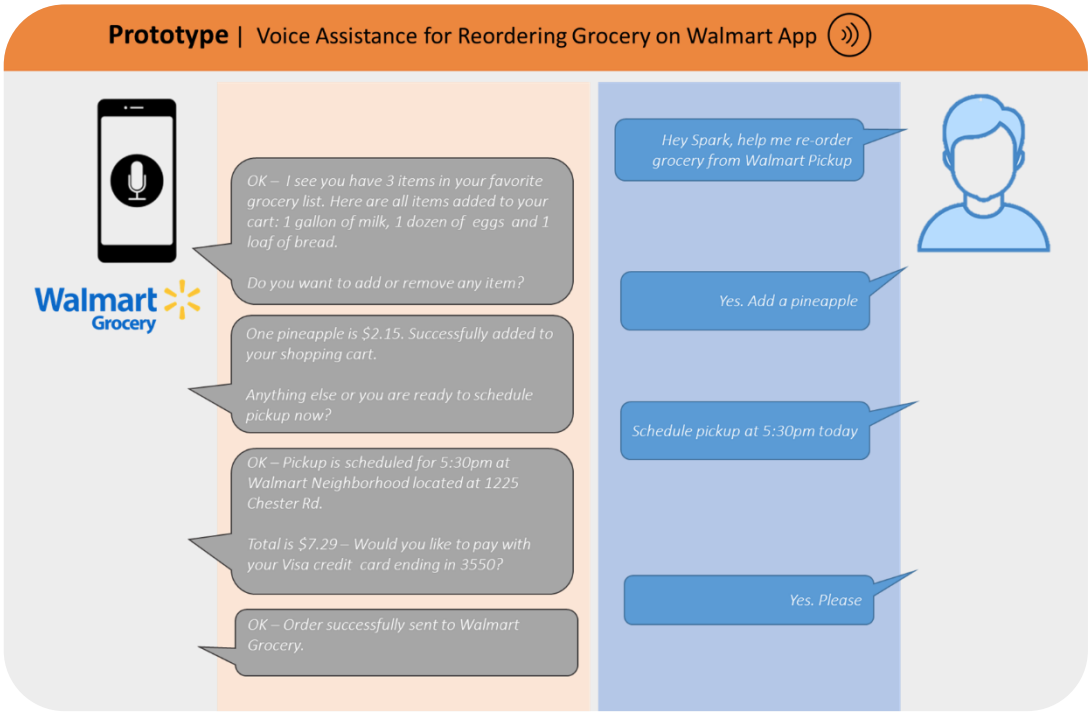
Appendix A

Prototype#2: Easy Auto Reorder Interface



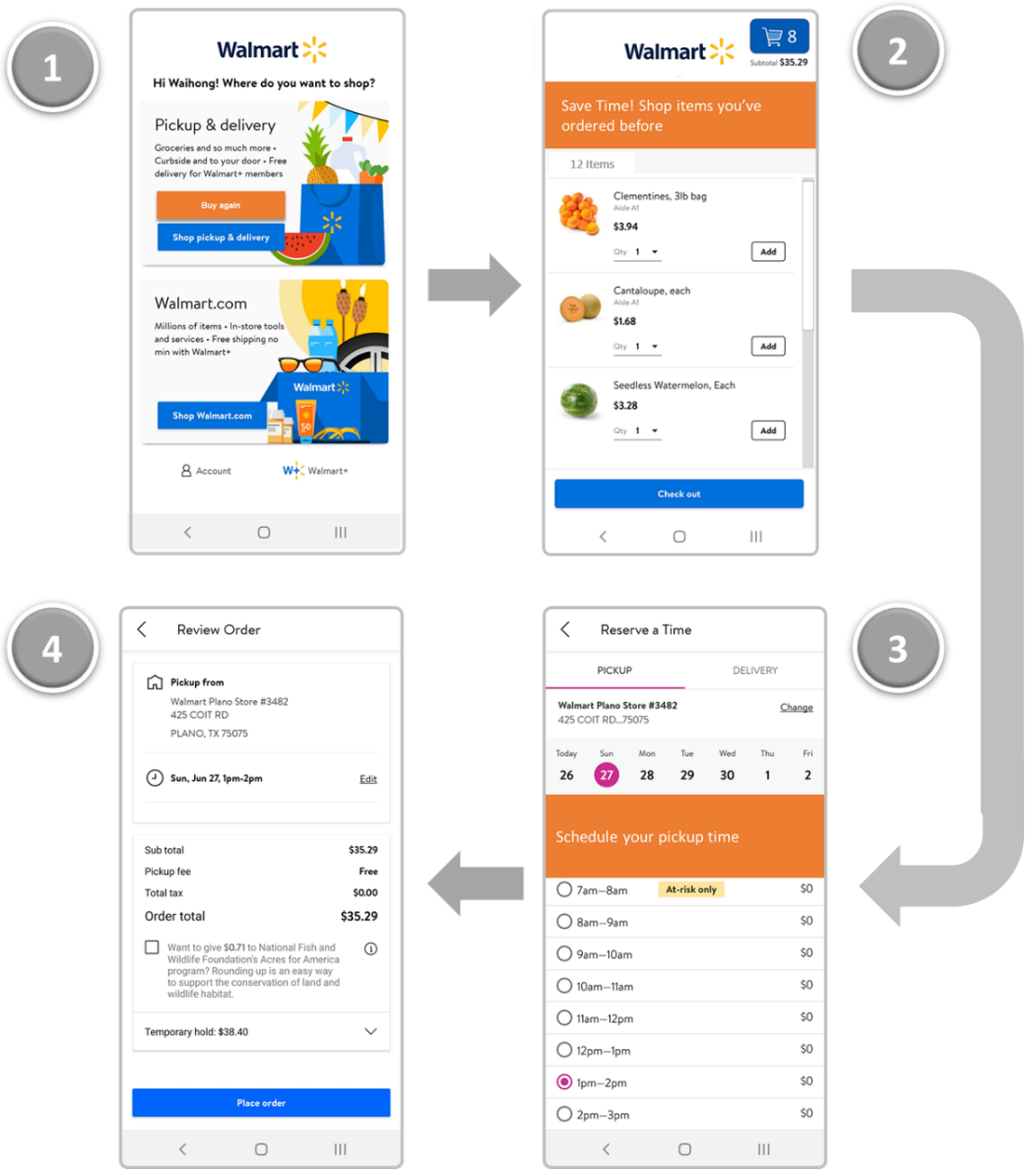
Appendix B

Prototype#1: Voice Interface for Walmart App



Appendix C

Prototype#3: Buy Again Interface



Appendix D

Table 1 – Requirements for Walmart app redesign

	Design Requirements
Functionality	○ Primary function shall allow the users to create a frequent or favorite grocery items list that can be easily accessed within the primary interface. ○ New items shall be added to the list directly from the browsing page.
Learnability	○ The feature shall be discoverable and easy to learn for novice users. ○ The design of buttons shall meet the principle of affordances.
Usability	○ The design shall reduce the time and effort required for users to re-order grocery items, with a maximum target of not to exceed 2 minutes.
Reducing Cognitive load	○ The interface shall be simplified to reduce complexity ○ Provide direct access to re-order items (shopping list) with simple process flow

Table 2 – Data Inventory for Walmart app redesign

Data Inventory Item	Insights
Who are the users?	Based on the demographic information collected from the survey, we learned that the users are middle age individuals, married with at least 1 child in the household. Younger adults who are single with no kids, are less likely to shop grocery online. Another group of probable users are the retired seniors (age 65+). However, we were not able to validate this users group since the survey was sent to our fellow students who are outside of this age range.
Where are the users?	The users can be from any geographical location, but service availability for pickup and delivery may limit how far the users live within the area

Data Inventory Item	Insights
	of a nearby Walmart store. Because of this reason, we learned that *Instant cart* has an advantage over Walmart app due to their ability to service a wider range of local grocery stores.
What is the context of the task?	The users predominantly complete this task from home or from their offices using the Walmart mobile app. The users are often very busy taking care of the family and most likely to multitask while shopping grocery online.
What are their goals?	Primary goal is to obtain grocery and necessity items for the family.
What do they need?	Saving time is important to the users. When it comes to grocery shopping, they are looking for convenience, and ways that can help maximize the efficiency of completing this routine activity. We learned that the Walmart app today have failed to satisfy their needs. The interface is overly complicated and difficult to use.
What are their tasks?	The main task here is to re-order frequent grocery items from a shopping list or prior purchases for a quick checkout process.
What are their subtasks?	The subtasks necessary to complete the above task are as follow: - Create a new list - Search or browse items to be added to the shopping list - Transfer items from shopping list to basket - Initiate checkout - Schedule pickup location and time - Complete payment - Send order