Sprouts Sign-Ups: Designing an Interface to Reduce Information Overload and Maximize Scheduling Performance

Introduction (5pts)

Sprouts is a non-profit, volunteer-run organization at UBC that seeks to provide affordable and sustainable food for both UBC students and the local community. The program, which initially started with a small number of volunteers, has grown in size to nearly 250 people. Coordinating shifts, signing up for initiatives and organizing volunteers are just a few of the goals Sprouts members are tasked with, with no single interface satisfying the team's needs. Therefore, the scheduling tactics originally used by the organization have become inadequate, due to the sheer amounts of volunteer data that must be organized.

Given these issues, we sought to address coordination difficulties in Sprouts by designing an interface for volunteers to provide shift availability. We designed two prototype interfaces: a single calendar interface (with all initiatives and shifts presented in one calendar) and a multiple calendar interface (with shifts for different initiatives divided between multiple tabs), as field study data indicated that having a more visually coherent interface may help lessen the burden of information overload. Additionally, each interface had a variant with a 'selected shifts' box to test the value of additional visual feedback during sign-up. We evaluated both interface efficiency and user satisfaction post-interaction in hopes of determining a design direction assisting Sprouts in conducting their day-to-day scheduling activities.

Experiment description (with divergences from MSIII clearly marked) (8pts)

Goals

The first goal of our experiment was to determine which interface, if any, maximized the satisfaction when conducting scheduling tasks for Sprouts volunteers. Satisfaction was measured via post-task surveys including Likert scales and ranking of interface designs.

The second goal of the experiment was to determine which interface, if any, maximized the efficiency when conducting scheduling tasks for Sprouts volunteers. Efficiency was measured via the time it took for users to complete their scheduling tasks.

The final goal was to determine whether increased redundancy in scheduling information, implemented via a 'shifts selected' box, would impact either of the above factors of satisfaction or effectiveness.

Participants
We recruited N=12 UBC Sprouts volunteers via convenience sampling via a group member who was already a member of Sprouts. The ages of the individuals ranged between 18-23, with the average age being 20.6. The participants held a variety of roles, including 2 volunteer coordinators, 4 senior volunteers, and 4 junior volunteers. The remaining 2 participants were UBC students not in Sprouts.

Conditions
There were both within-group and between-group conditions presented to participants. The within-group treatment consisted of varying interface types, with participants using three different interfaces: a single calendar interface (with all initiatives and shifts presented in one calendar view), a multiple calendar interface (with shifts for different initiatives divided between multiple tabs), and the current sprouts scheduling form which acted as our control. The between-group treatment consisted of either the presence or the absence of a 'shifts selected' box, a UI element that summarized all the shifts that the user had selected.

Experimental Tasks
Participants were asked to sign up for 2 shifts in each of the 4 initiatives (for a total of 8 shifts). We standardized the shifts they sign up for as follows:

- Cafe: Monday 11:00AM–1:00PM, Thursday 3:00-5:00PM
- Community Fridge: Tuesday 6:00-8:00PM, Wednesday 1:30-3:30PM
- Distro: Saturday 10:00AM-12:00PM, 2:00-4:00PM
- Community eats: Friday 9:00-11:00AM, 2:00-4:00PM

Apparatus
The experiment was conducted in-person, with the participant seated at a desk with the experimenter positioned behind them during the experiment. Participants completed the experiment using the experimenter's computer (either a Macbook Pro or Microsoft Surface Pro).

Design
The experiment was a split-plot factorial design, with the aforementioned factors:

- **Between-groups:** Presence of shifts selected box
- **Within-groups:** Calendar view

Time to complete the task was measured as the time from when the experimenter told the participant to start to when the user had clicked the submit availability button. User

satisfaction was measured measured using a post-experiment questionnaire assessing the following items:

- Ranked preference between all 3 calendar types
- Aspects of the interface they liked
- Ease of use of the interface
- Usefulness of the interface

Other than the ranking question, all questions used Likert scales to quantify data. Some of the survey questions were adapted from the User Experience Questionnaire (UEQ; https://www.ueq-online.org/).

Procedure

1. Before the experiment we randomly assigned participants to the between groups condition (shift selected box present vs. not present), and counterbalanced the within groups condition (i.e., the order in which the interface types were presented).

2. The med-fi prototypes were prepared based on the participant's assigned conditions. Note that for the control, we started the task from the 2nd page of the form (after the page where the participant filled in their details).

3. Briefed the participant and obtained written consent

4. Began the timer and began tasks with the participant.

5. Stopped timer once tasks were complete.

6. Administered post-experiment questionnaire, debriefed participant and ended experiment.

Independent Variables

- Interface type (current form, single calendar, multiple calendars split by initiative)
- Presence of 'selected shifts box'

Dependent variables

- Time to complete the task
- User satisfaction

Hypotheses

H1: Users would have higher satisfaction with at least one of the tested alternatives than with the original google form method of submitting availability.

H2: Users would be faster in completing the task with the split view calendar than with the single view calendar.

H3: Users would be faster in completing the task with the ‘without selected shifts box’ than with the ‘with selected shifts box’ interface.

H4: There would be a difference in the time to complete the task for the split view calendar and the single view calendar for each selected-shifts box interface (i.e., an interaction between calendar view and shift selection interface).

Statistical analyses

Experimental Task Data: Split-plot ANOVA, followed up with post-hoc t-test analyses to determine which groups are significantly different from each other.

Survey Data: Responses were quantified (since we used rankings and Likert scales), then subjected to a Friedman test to determine whether there was a significant difference in any of the ratings. We then followed up with post-hoc Wilcoxon signed-rank tests.

Results (including Appendix A.I - supplementary analysis) (12pts)

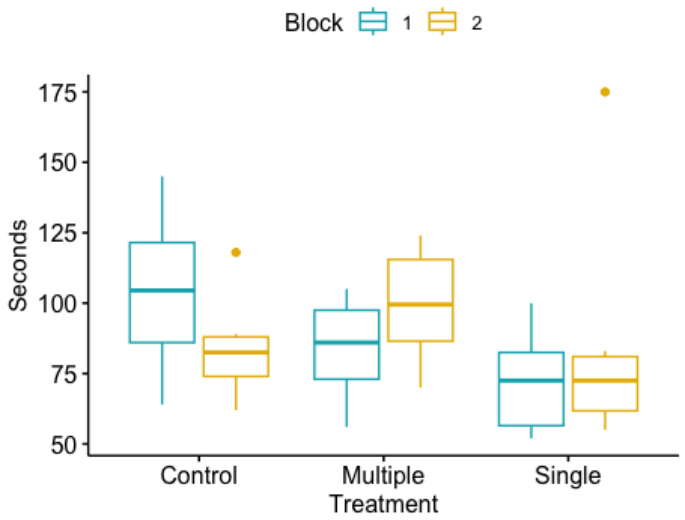
Efficiency Task
To test for a main effect of calendar type (Treatment), a main effect of selected-shifts box (With Block 1 being box is absent, and Block 2 being box is present), and/or an interaction between the two, we ran an ANOVA test with our sample size of N=12.

We had two outliers: one in control for Block 2, and one in single calendar view for Block 2.

Results: No statistically significant results in terms of main effect of calendar view (P-value=0.3634918), main effect of selected shifts box (P-value=0.7320387) or interaction effect with both (P-value=0.3602843). See Figures 1, 2 and 3 in Appendix A1 for summarized results.

However, trends-wise we see that the single view calendar was the fastest for both blocks (selected shift box and no selected shift box), and the multiple view calendar was faster in Block 1 versus control, but control was faster in Block 2 versus multiple view calendar (see Figure 1 for visual display of trends).

Figure 1. Plotting task time with respect to Treatment and Block

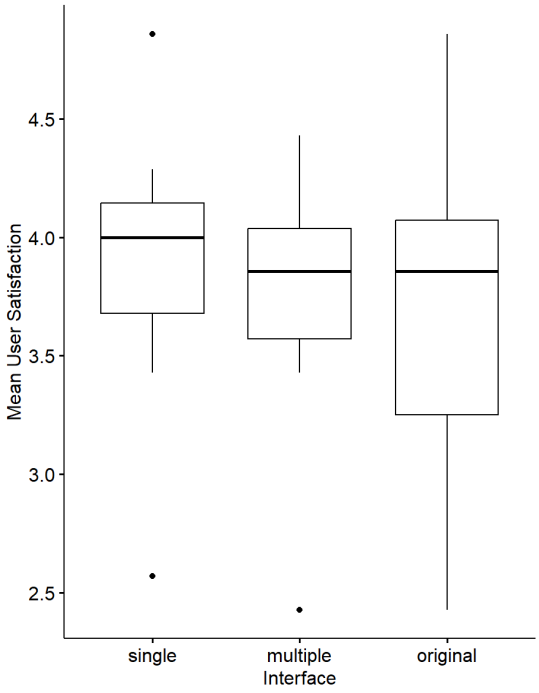


The post-hoc analysis tells us that, with the previously mentioned trends, using multiple calendars view resulted in completing the given task ~3 seconds faster compared to control, and the single calendar view resulted in completing it 15 seconds faster. The single calendar view compared to the multiple calendars view was ~12 seconds faster.

Note that we conducted a post-hoc test to further explore trends in the data despite there being no statistical significance. The Tukey test was used due to its functionality being that of a standard t-test, while also correcting for family-wise error rate (the probability of at least 1 false positive occurring when multiple comparisons are being tested).

User Satisfaction:
To test for any significant difference in user satisfaction across any of the interfaces, we collapsed the satisfaction questions from the survey into a single mean satisfaction value (this method is consistent with *Findlater, Moffatt, McGrenere, and Dawson, 2009*). Because the data was collected on Likert scales, it is non-parametric. Therefore, we used a Friedman test.

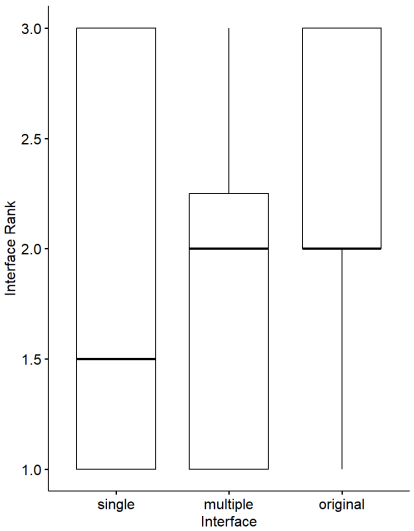
Results: no significant effect of interface on mean user satisfaction (P-value=0.6616). See Figure 5 in Appendix A1 for summarized results.



No post-hoc tests were necessary, as no results were significant. Further, the primary analysis did not seem to suggest any trends that warranted further analysis.

Interface Rankings:
As part of the post-interaction survey, we asked participants to rank the interfaces from most-to-least favoured. We performed another Friedman test to determine whether there were any significant differences in the rankings across interfaces.

Results: No significant difference in ranking across interfaces (P-value=0.558). See Figure 6 in Appendix A1 for summarized results.



Post-hoc Tests: Although the results were not significant, we noted that the rankings of the single calendar interface seemed like a trend and therefore warranted further investigation. We compiled descriptive stats on the data as a post-hoc followup to determine any trends in the data.

Post-hoc Results: half of participants (6/12) ranked the single calendar interface as their most preferred interface. Overall, 10/12 (83.3%) of participants ranked one of the prototypes over the original interface. See Figure 7 in Appendix A1 for summarized results.

Step 4: Identify limitations to the experiment (individual)

Given the results of the experiment, we see support for Hypothesis 1. That is, users had higher satisfaction with our alternative interfaces as opposed to control. However, despite the interesting trends we see in our data, several limitations are present in our experiment that threaten the validity of our results.

Firstly, there was an uneven representation of the various Sprouts roles due to challenges in recruiting participants, with a majority of participants having more senior/coordinating roles. This means that our population is not necessarily representative of our target demographic and thus our results less generalizable to the Sprouts community as a whole, such as with incoming or Junior volunteers.

Another issue was that of methodological inconsistencies. For some participants, particularly with the first few runs of the experiment while we were getting familiar with the protocol, we failed to explain the tabbed versus non tabbed interface designs. Going forward, we began giving more in-depth explanations of how the prototype interfaces functioned to reduce negative transfer of learning. However, this highlights one potential confounding variable we encountered: that of participant familiarity with the control interface, resulting in a significant decrease in time to complete tasks with control treatment. An additional methodological pitfall was that during task completion, we asked participants to sign up for the “Fridge” initiative, which was labeled accordingly on our prototypes. However, this was called “Donation Restock” on the control interface, which likely biased participants towards having increased ease of sign up with our interfaces for the given category.

Looking at our results post-experiment, although we see trends of improved speed and satisfaction with our interfaces, one large methodological issue was that of the number of initiatives. The control form had a greater number of initiatives (5) while our interfaces only displayed 4. This fifth initiative category, that of Sprouts Driver, also included other additional information present in the control such as questions asking volunteers about

having a driver's license or having previous experience being a Driver. Since our prototype was not wholly reflective of the actual information present on a typical Sprouts form, we cannot say for certain that our results are valid, as the extra information adds to time spent on task and possibly impacts satisfaction as well. Testing our interfaces with the accurate amount of information/initiatives is essential in future work, in order to examine possible changes such as in satisfaction and speed, or cognitive overload.

Regarding the remaining hypotheses, we did not see support for Hypothesis 2. On the contrary, instead of seeing faster task completion with the split calendar view, we saw faster task completion using the single calendar display. One possible reason for this outcome stems from a persisting comment seen by several participants, being that the tab bar on the multiple-calendar view was too far from the actual shift sign ups and was thus hard to initially locate. This is a design flaw that should be addressed in future iterations. A more general rationale is that there was simply not enough data to warrant having shift times separated into multiple views. It may be that there is some threshold in terms of data saturation that makes a tabbed calendar superior to a single-view calendar, and that the current number of initiatives does not yet reach that threshold. For example, we might expect participants to increasingly prefer the tabbed calendar as initiatives grow in number and data becomes more quantitatively/visually dense.

There was also no support for Hypotheses 3 or 4, regarding either a main effect or interaction effect of the “selected shifts” component. Again, one may attribute this to an association between the amount of data presented and its impact on task speed/satisfaction. It would be interesting to run the experiment with a greater amount of data, to see if having the redundancy of the selected shifts box become of greater assistance as the amount of data increases.

Step 5: Formulate experiment conclusions (individual)

After testing two Sprouts scheduling prototypes for both efficiency and user satisfaction, given the current insights gained by our experiment, the single-calendar view of shift sign ups shows promise with regards to design direction. Although our results were not statistically significant, the superiority of the single-calendar interface can be largely attributed to a combination of qualitative and quantitative results. Trends-wise, a large majority of participants preferred the calendar prototypes over the control, and the single-calendar view proved to have a substantial decrease in time for task completion.

This outcome, in favor of the single-calendar view, was a surprise to our team, as we suspected that the tabbed view would decrease cognitive overload on users and subsequently improve both task speed and satisfaction. However, it seems that users liked having all shift information available to them on one page, saving the hassle of

switching back and forth between tabs. Additional positive features of the single-calendar interface included the more complex visual representation of shift times, such as the use of different colours/labeling to differentiate the different available initiatives.

The selected shifts box did not seem to have an effect on task performance, which also proved surprising to the team, with neither a main nor interactive effect seen in analysis. Again this brings into question whether there is a threshold of data saturation that may impact any of our findings, and also highlights that, with all of our results in general, there are limitations that likely affected our results.

Team Sprouts will continue working on the design once the term ends as a self-directed project. Becoming more familiar with the Sprouts program, meeting the coordinators, and becoming better acquainted with the volunteers gave personal meaning to the project and motivates the completion of a scheduling interface that better supports the organization.

Appendix A1

Figure 1. ANOVA Testing main effect of Treatment (P-value=0.3634918):

```
            Df Sum Sq Mean Sq F value Pr(>F)
Treatment    2   1537   768.4   1.044  0.363
Residuals   33  24297   736.3
```

Figure 2. ANOVA Testing main effect of Block (P-value=0.7320387):

```
            Df Sum Sq Mean Sq F value Pr(>F)
Block        1     90    90.2   0.119  0.732
Residuals   34  25744   757.2
```

Figure 3. ANOVA Testing interaction effect of Treatment x Block (P-value=0.3734839):

```
                Df Sum Sq Mean Sq F value Pr(>F)
Treatment        2   1537   768.4   1.056  0.360
Block            1     90    90.3   0.124  0.727
Treatment:Block  2   2385  1192.3   1.639  0.211
Residuals       30  21822   727.4
```

Figure 4. Post-hoc Analysis of Treatment + Block

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Seconds ~ Treatment + Block, data = data)

$Treatment
                       diff       lwr      upr     p adj
Multiple-Control  -2.666667 -30.25924 24.92591 0.9694155
Single-Control   -15.000000 -42.59257 12.59257 0.3861939
Single-Multiple  -12.333333 -39.92591 15.25924 0.5220064

$Block
        diff       lwr      upr     p adj
2-1 3.166667 -15.50799 21.84132 0.7320514
```

Figure 5. Friedman Rank Sum Test of User Satisfaction

```
        Friedman rank sum test

data:  data$mean, data$interface and data$id
Friedman chi-squared = 0.82609, df = 2, p-value = 0.6616
```

Figure 6. Friedman Rank Sum Test Across Interfaces

```
        Friedman rank sum test

data:  rankings.long$rank, rankings.long$interface and rankings.long$id
Friedman chi-squared = 1.1667, df = 2, p-value = 0.558
```

Figure 7. Post-hoc Analysis of Interface Preferences

```
rankings$single
Type: Integer

              Freq   % Valid   
----------- ------ --------- --
          1      6     50.00   
          2      2     16.67   
          3      4     33.33   
       <NA>      0             
      Total     12    100.00   

rankings$multiple
Type: Integer

              Freq   % Valid   
----------- ------ --------- --
          1      4     33.33   
          2      5     41.67   
          3      3     25.00   
       <NA>      0             
      Total     12    100.00   

rankings$original
Type: Integer

              Freq   % Valid   
----------- ------ --------- --
          1      2     16.67   
          2      5     41.67   
          3      5     41.67   
       <NA>      0             
      Total     12    100.00
```

Pilot Study

A number of changes emerged after we completed our pilot that broadly fell into two categories: 1) methodological issues that dealt with the design of our experiment and materials; 2) logistical issues that dealt with the practicalities of actually carrying out this study.

Methodological

Prototype feedback: As our prototype was still WIP at the time of the pilots, we received some small feedback around certain elements. For example, the shift sign-up displayed a specific date and time. The shift selection was also not properly updating the styling of an event after it had been signed up.

Questionnaire: We pared down the number of Likert-type items in our questionnaire because some of them were perceived as too similar. For example, “Is practical and useful” and “does what I need it to do” were thought to be interchangeable, so we kept the latter as it more plainly communicated what we wanted participants to consider.

Task changes: To make the task more realistic, we added all the shift times for each initiative that were present in the original interface. We had initially received complaints in our field study that information overload was an issue with some of the current state systems, so we sought to adequately reflect this in our two prototypes.

Accounting for learning/habituation: We noticed that our pilot participants had some initial trouble navigating the prototypes and getting used to the task. We thus made the following changes:

1. Include in our procedure some time to show participants the interface before beginning the task
2. Ask participants what their preferred trackpad orientation was and adjust it on the experimenter’s computer

We hoped that these changes would reduce the impact that trying to learn how the interface works would have on our tasks (as we were interested in task completion time). While having a number of practice trials would have been more ideal, we decided to go with this approach to balance practical constraints with recruitment and the number of participants we were aiming to run.

Logistical

We made the following changes based on difficulties we encountered administering the experiment in our pilot:

1. Capture screen recordings to take timestamps instead of timing by hand in order to be 1) more exact and 2) have less things for the experimenter to worry about
2. Prepare adequate printouts ahead of time. In particular, we aimed to provide the list of shift times the participant had to sign up for ahead of time.

3. Clearly specific in the protocol that participants should wait for us to tell them to start (as one of our pilotees jumped immediately into the task once the interface was shown)
4. Clearly specify the stopping action (e.g., press submit, navigate to end of form) as one of our pilotees simply waited around without being sure what to do.

References

Findlater, L., Moffatt, K., McGrenere, J., & Dawson, J. (2009). Ephemeral adaptation: The use of gradual onset to improve menu selection performance. Paper presented at the 1655-1664. https://doi.org/10.1145/1518701.1518956