CPSC 444: Milestone IV - Individual Report (MSIV)

Naomi Tran
April 12, 2023

Introduction

Physical inactivity is prevalent during young adulthood due to lack of time, motivation, and accessible places (Silva et al., 2022). With less time, people become more overwhelmed and can become burnt-out by all the responsibilities they need to do. Besides dropping physical activity, adults find making plans with their friends more difficult due to having to go back and forth to find a time that aligns with everyone's schedules (Latham, 2019). Although there are tools that aim to solve these problems, there is not one that makes it easier for people to find friends to do physical activities with. To discover more about this issue and the barriers that exist in people's lives that prevent them from exercising, we conducted a field study with six participants over Zoom. We interviewed them to learn about their difficulties when it comes to being active and what motivates them to be active. We also wanted to understand how they schedule plans with their friends and what platforms they currently use to support these needs.

From our field study, we learned that users are more motivated to be active when they have friends to do it with because they feel a sense of camaraderie, but it can be difficult to schedule. This led to the creation of *Find Fit Friends*, which helps people find friends to do physical activity and facilitates finding a time that aligns with everyone's schedules. Although this study does not focus on the algorithm behind our scheduling feature, we wanted to evaluate the user interface design since there is not a platform that currently allows users to schedule activities with people they meet on an app.

Specifically, this study evaluates the effectiveness and intuitiveness of our scheduling feature on how well our two prototypes support four specific scheduling tasks. Thus, we conducted a two-way experiment with six participants, collecting qualitative and quantitative data from our tasks and follow-up interview. We piloted the experiment with two participants and updated the experiment and prototypes accordingly.

Goals

The first goal was to determine which prototype design was more effective for users when looking through their calendar. To measure effectiveness, we looked at the time it took for users to complete a specific task. The task created was centered around the task examples and field study findings. Since scheduling revolves around planning and looking through dates, finding an activity is an important task.

The second goal was to determine whether a prototype design supported editing and/or creating an activity better than the other design. Since editing and creating activities share many similar steps, we wanted to see whether users are affected by any minor differences between the two designs. To measure this, we looked at the number of errors, as this would reflect the intuitiveness and/or learnability of the prototypes.

The last goal was to determine which design was more intuitive based on their subjective ratings when looking through invitations. Since the prototypes have completely different designs, we wanted to know whether one design supports users better than the other.

Methods

Participants

We recruited 6 participants (3 male, 3 female) using convenience sampling by reaching out to friends and acquaintances. Our participants were university students aged 18 to 24 with varying fitness levels and goals, and diverse backgrounds and experience to reflect our target users like our field study. Since we developed our prototypes based on findings from our field study's participants, we were confident this group would provide valuable feedback to help us evaluate the usability of our prototypes for our target population.

Conditions (see Appendix A.II for prototypes)

We have two prototypes: *Weekly* (Figure 1a) vs *Monthly* (Figure 1b) view.

Weekly. This prototype features a weekly calendar view with a horizontal scroll to look through other weeks. To edit an already booked / create an event, this prototype breaks up the steps between inviting a user and the other steps. It also has a more visual location option and an open text field to input the activity type, but requires users to scroll through the page to complete these steps. Moreover, the invitations on this prototype are all on the primary calendar screen.

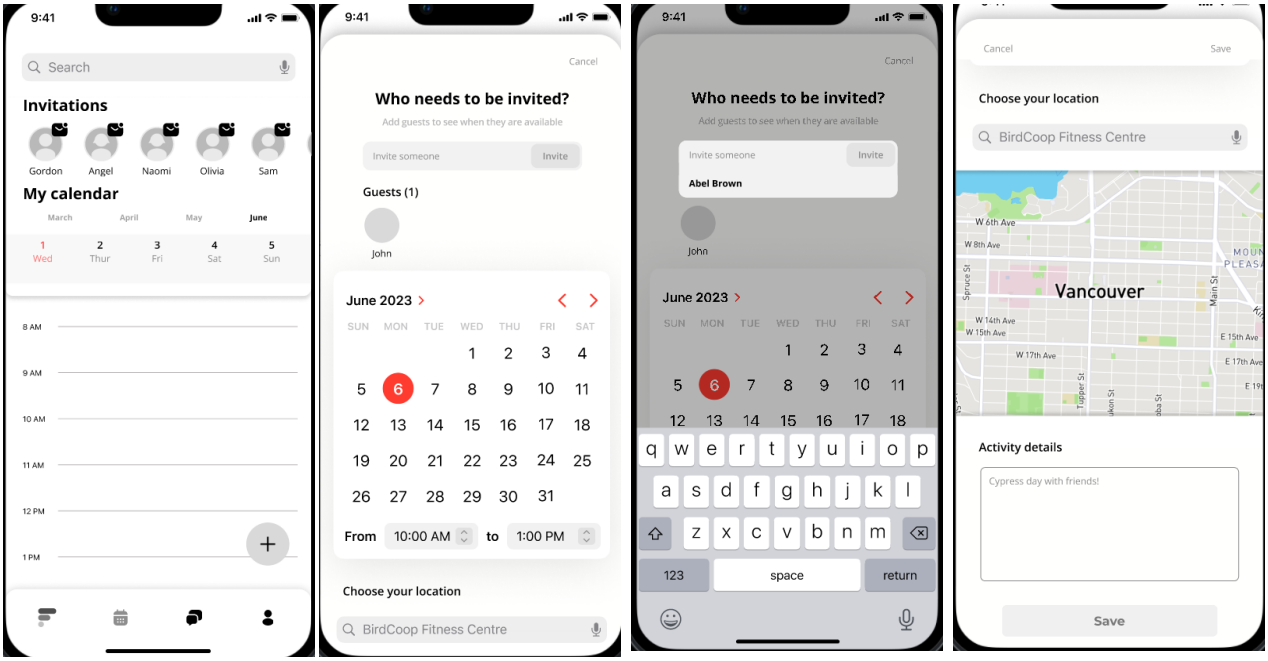


Figure 1a: Screenshots of *Weekly* prototype; calendar (including invitation display) and editing an activity page (including how one invites a guest and the displays for location and activity).

Monthly. This prototype features a monthly view similar to Apple and Google calendar. To edit an already booked / create an event, this prototype has a single page for editing details like activity type and invitees. The user can only select one of the predefined activity types while the location option is an open text field. The editing / creating event feature fits on one page, without needing to scroll. The invitations on this prototype are housed in an "inbox", requiring users to click on the inbox icon, taking them to a separate page with the invitation features (look through, accept / decline).

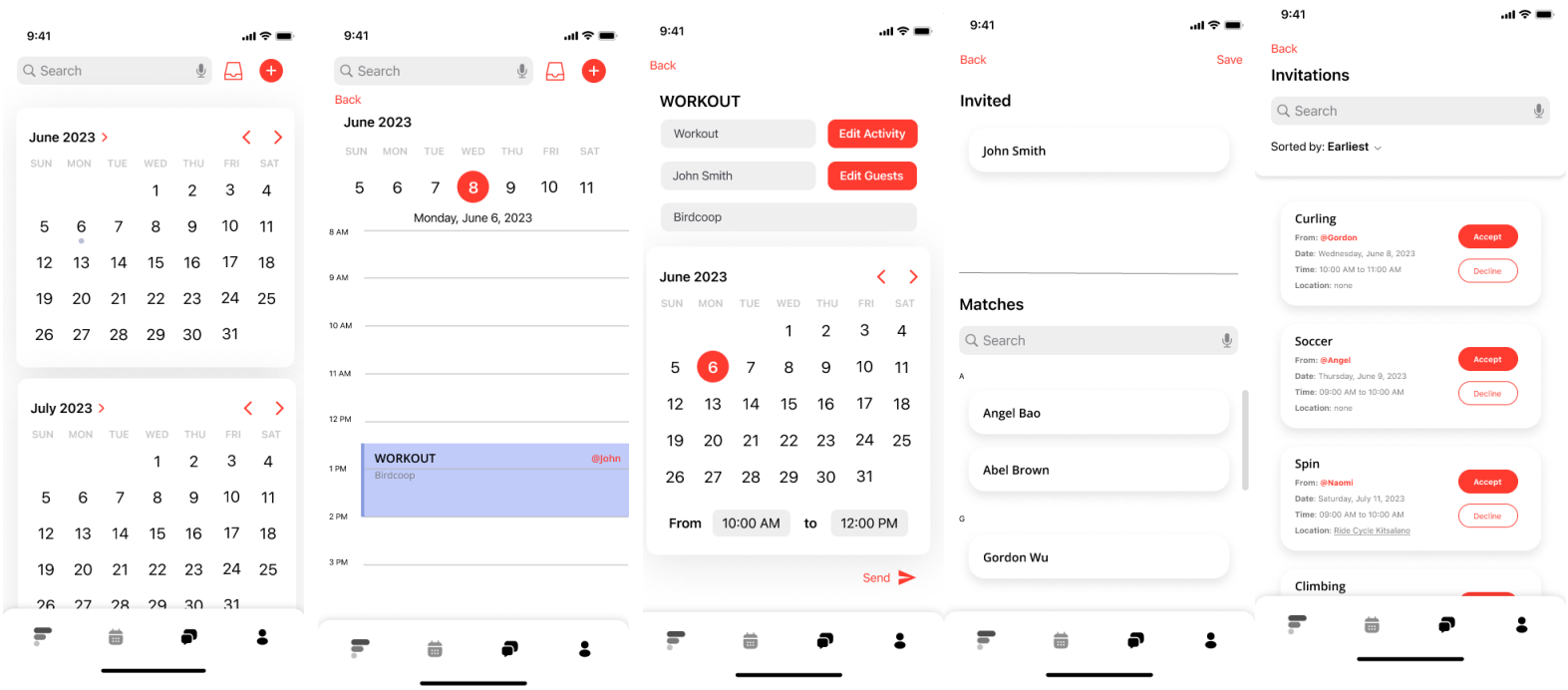


Figure 1b: Screenshots of *Monthly* prototype; calendar (with inbox for invitations), editing an activity page, inviting guests page (same display for choosing an activity), and invitation page.

Experimental Tasks

Participants had to complete the same four scheduling tasks in the same order for both prototypes (see Appendix A.II for task details).

Task 1: Find an already booked event within the calendar tab.

Task 2: Edit an already booked event by adding another guest and changing the date and time.

Task 3: Create and send a new event to a connection.

Task 4: Look through their pending invites to accept a specific invite.

Design

Our experiment was a 2x4 factorial design with participants completing 4 tasks (within-subjects) on 2 prototype designs (*Weekly*, *Monthly*; within-subject). We counterbalanced the order of the prototypes used by having half of our participants complete tasks with *Weekly* first and the other half using the *Monthly* first to avoid carryover / learning effects. We used a remote observation over Zoom and gathered any comments they made and their behaviour during their tasks to help us identify areas of confusion. We also collected the time it took to complete task 1 and the number of errors for task 2 and 3. After participants completed all four tasks on one prototype, we conducted a semi-structured interview to understand how intuitive they thought it was to complete each task and why. We repeated the same process for the next prototype.

Procedure

First, the experimenter explained the experiment's purpose and overview and obtained consent for audio and screen recording**.** After answering any questions the participant had, they were given the first prototype to open up on Figma. Odd numbered participants received *Weekly* first, and even numbered participants received *Monthly* first. Once participants were ready, they were given each task one by one, while the experimenters noted down observations on the coding sheet (see coding sheet in Appendix A.II). For task 1, experimenters also recorded the time it took to complete the task. For tasks 2 and 3, experimenters recorded the number of incorrect clicks and noted when participants showed behaviour that reflected confusion (i.e. mouse hovering, head tilt, furrowed eyebrows, etc.). When participants seemed confused or made comments like "I think this is frozen" or "I think this is broken", experimenters guided them towards the correct step, and noted down where this happened.

After completing the tasks with their first prototype, the experimenter conducted a short interview focused on their experience for that specific prototype (see interview questions in Appendix A.II). This interview included Likert scale-type questions to get their level of intuitiveness and open-ended questions to encourage participants to reflect and give their feedback and thoughts on their experience. Then, participants were given the other prototype to complete the same tasks, followed by the same interview questions.

Once all the tasks on both prototypes were completed, experimenters asked whether participants had any questions before thanking them for their time.

Apparatus

The experiment took place over Zoom where participants could choose a comfortable environment with minimal noise to prevent any distractions and use their own desktop devices. Zoom was used to screen and audio record the experiment so we could review and analyze the data later, providing us with more accurate and detailed insights. Zoom's screen sharing feature also allowed us to observe the participants' actions and comments as they completed the tasks using the prototypes.

A stopwatch was used to measure the time it took for each participant to complete task 1 for both prototypes. A spreadsheet (coding sheet) was used to record the time, number of errors, intuitiveness, and observations.

Independent and Dependent Variables

Dependent Variables

Time. We measured the time it took to complete task 1. We used a stopwatch and/or analyzed our screen recordings. This helped us determine the effectiveness for looking through the calendars on our prototypes; a faster time suggests the prototype is more effective.

Number of errors. We measured the number of errors it took for tasks 2 and 3. This was based on the number of incorrect clicks. We counted multiple clicks in the same area at the same time as one incorrect click, but if participants went back to click on it again, this was counted again. This helped us determine the intuitiveness for our prototypes when editing / creating an activity in their schedule; less clicks when completing a task suggests the prototype is more intuitive / easier to learn.

Intuitiveness subjective ratings. We asked users to state their level of intuitiveness when completing task 4 on a scale from 1 to 10 (1 = very difficult; 10 = very easy). This Likert scale question helped us determine the intuitiveness for our prototypes when accepting an invitation; a prototype with a higher rating suggests it is more intuitive.

Independent Variables

Prototype. We have two prototypes – *Weekly* and *Monthly*. Every participant used both prototypes so we could compare the differences in prototypes, and determine which one was more effective and intuitive.

Tasks. We have four scheduling tasks. Every participant completed all four tasks twice (the same tasks on both prototypes).

Hypotheses

To evaluate which prototype supports scheduling tasks better, we looked at three metrics – speed, accuracy, and subjective ratings. Speed was our quantifiable metric for evaluating effectiveness; accuracy and subjective ratings were our quantifiable metrics for evaluating intuitiveness / learnability.

H1: Users will be faster at finding an already booked activity (task 1) with the *Monthly* prototype.

H2: The performance of the prototype design on the number of errors changes depending on editing or creating an event (task 2 or 3 respectively), or vice versa.

H3: Users will find the *Weekly* prototype to be more intuitive when looking through pending invitations to accept a particular one (task 4).

Planned Statistical Analysis

There are three factors that we measured in this experiment: time to complete task 1, number of errors on task 2 and 3, and intuitiveness on task 4. Due to our experimental design, we used a t-test for task 1 and 4 to answer H1 and H3, as this allowed us to compare the means between prototype designs on the specific task to determine which one is faster (H1) or more intuitive (H2). We used a two-way ANOVA to answer H2 as this allowed us to see how the mean of our dependent variable, number of errors, changed according to our two independent variables, prototype design and task.

Limitations

We had a small sample size of six participants due to our time and resource constraints. This limited the amount of data we could collect and analyze, making it difficult to determine whether our findings are true. Specifically, we are more prone to make a type II error, incorrectly accepting the null hypothesis when in fact there is a difference between the study groups.

Additionally, the time to complete task 1 and the number of errors for tasks 2 and 3 were collected by four different evaluators. Having four different perspectives may have led to subjective and varying beliefs as to what was considered “complete” or an “error”. Although we had discussed exactly when to start and stop the timer, and when to count a click as an error, these nuances could have led to inconsistent findings due to human error.

We also recognize the nuances that come with using a mobile versus a desktop device. Although our prototype is supposed to be used as a mobile app, time and resource constraints prevented us from conducting the experiment with participants in real life on a mobile device, which could have caused more or less issues with certain tasks. Instead, we used Figma prototypes on desktop via Zoom, but this made it difficult to actually measure what we thought we were measuring since scrolling and/or other touchscreen actions on a mobile device is much different than dragging on a mouse or touchpad. If users were to use mobile devices, it may be more obvious that the screen is scrollable, as any tiny movement would move the screen.

Finally, both prototypes lacked some functionality due to time and resource constraints. We had designed our prototypes to follow specific steps in order to complete the tasks. This may have led to more “number of errors”, even when participants may have just wanted to do it in a different order (i.e. wanting to choose a date before inviting a guest). Furthermore, our *search bar* offered no real affordance as we did not completely finish designing this feature since we were not evaluating this. However, this may also have caused more “number of errors” since participants were likely to use this function to search more easily.

Results

Quantitative (see Appendix A.III for raw data)

	Task 1 (time: sec)	**Task 2 (# errors)**	**Task 3 (# errors)**	**Task 4 (intuitive: x/10)**
A	32	2.5	1.5	8.3
B	4	1	1.5	7.9

Figure 2: Mean findings for each task among all six participants for each prototype (A: *Weekly*; B: *Monthly*).

Time (task 1). From figure 2, the mean time for the *Weekly* prototype was 32 seconds with a maximum of 62 seconds and a minimum of 10 seconds. Two participants completed task 1 in slightly more than 60 seconds, while only one participant completed it in 10 seconds. For the *Monthly* prototype, three participants completed task 1 in around 4 seconds, one participant completed in 10 seconds (maximum), and two participants completed in around 2 seconds (minimum). As seen in figure 3a, the paired samples t-test resulted in $t(5) = 2.4174$, $p = 0.06031$. Although close, this was not statistically significant given a confidence interval of 95%, and thus, there was no main effect on the prototype design.

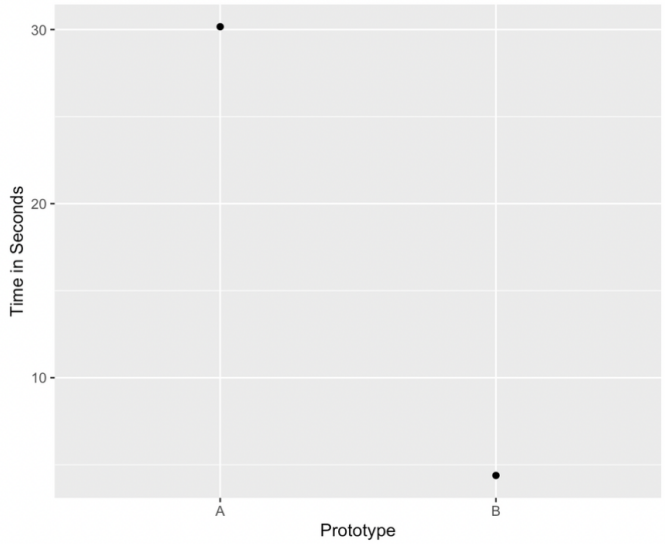


Figure 3a: This plot shows the t-test results for task 1's mean time for each prototype (A: *Weekly*; B: *Monthly*).

Number of errors (tasks 2 & 3). From figure 2, the average number of errors for task 2 for *Weekly* was slightly more than *Monthly* at 2.5 versus 1 respectively, but for task 3, they were identical at 1.5 errors. For task 2, the *Weekly* prototype produced a maximum number of errors at 5, while only one participant made no mistakes. For *Monthly*, the maximum number of errors was 4, while four participants made no mistakes. For task 3, one participant made six mistakes

with most making zero to one mistake with *Weekly*, while most participants made one to two mistakes with *Monthly*. As seen in figure 3b, the average number of errors converge but do not cross. The ANOVA results reported $F(1, 10) = 0.808$, $p = 0.390$, $\eta 2 = 0.050$, indicating there was no significant effect of the independent variable on the dependent variable, $F(1, 10) = 0.808$, $p = 0.390$, ns, and thus, no interaction effect between prototype design and task.

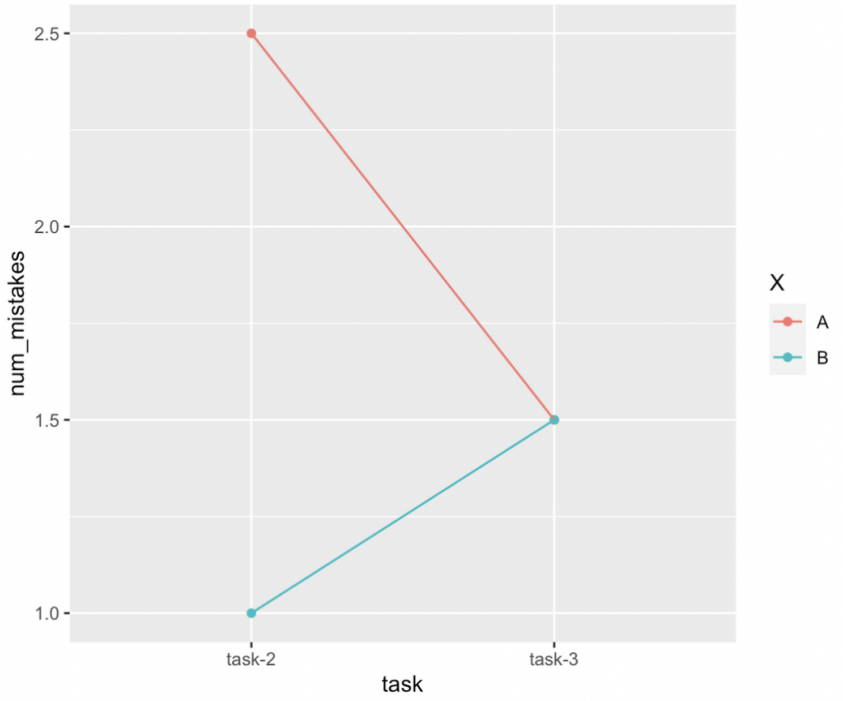


Figure 3b: This plot shows the repeated measures ANOVA test results for the number of errors made on tasks 2 and 3 for each prototype (A: *Weekly*; B: *Monthly*).

Intuitiveness (task 4). From figure 2, the *Weekly* prototype was slightly more intuitive to users at 8.3/10 while the *Monthly* prototype was 7.9/10. Two participants rated *Weekly* to have an intuitiveness of 10/10 (maximum) when accepting an invitation, with one participant rating it a 5/10 (minimum). Meanwhile, three participants rated *Monthly* to have an intuitiveness of 9/10 (maximum) and one participant rated it a 5/10 (minimum). In figure 3c, the paired samples t-test resulted in $t(5) = 1.5357$, $p = 0.1852$, indicating that this finding was not statistically significant with a confidence level of 95%, and thus, there appears to be no main effect on the prototype design.

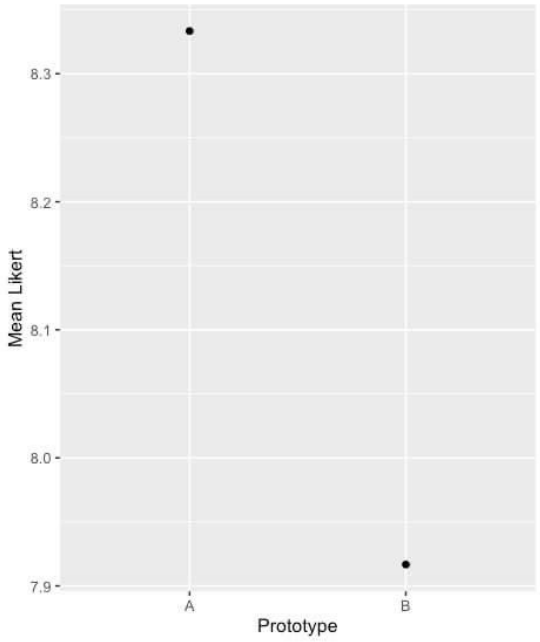


Figure 3c: This plot shows the t-test results for task 4's mean intuitiveness based on a Likert scale from 1-10 for each prototype (A: *Weekly*; B: *Monthly*).

Qualitative: Themes found in Weekly View Prototype (see Appendix A.I)

Scrolling issues. A unique issue to the *Weekly* prototype was that users had trouble scrolling vertically and horizontally, either due to being unaware of the affordance or not knowing how to scroll.

Pop-up issue. A unique issue to the *Monthly* prototype was that users had trouble closing pop-ups between task 2 and 3. Although we did not plan to evaluate the pop-ups, this issue prevented users from continuing the tasks and we counted these misclicks as an error as they completed task 3. Since the pop-up was grey and blended in with the interface, many completely missed it and struggled to close it before completing the next steps.

Confusion led to bottom-tab behaviour. When participants were confused on what they should do, they often redirected to the bottom tabs, and made comments like "Is it on another page?", suggesting they believed the steps they had to do were on a separate page.

Unaware. There were various cases when users completed a task but did not know and kept trying, and also thinking they have completed a task when they have not. For the *Weekly* prototype, a participant thought they had to find the calendar when they were already on it. Another participant thought they already invited a friend when they had not. For both prototypes, participants also thought the prototypes were broken when they were unable to click on certain buttons or scroll properly, when the design was not supposed to allow it. For

example, a participant tried clicking on the “invite” button before successfully choosing a guest to invite.

Mismatch in mental models. Both prototypes had multiple mismatches to users’ expectations. The steps users tried to take were not supported by these interfaces. For example, with the *Weekly* prototype, two users thought to click on “Angel” in the invitation list when asked to create an activity and invite Angel. Furthermore, participants did not expect the calendar to only display a week. Some explained how they were not used to seeing it like this, while others showed a number of errors when looking for an already booked activity like clicking on every day in the week to look for an activity, limiting themselves to only that week. And, we saw many participants hesitate or question when they had to fill out the activity type when creating an activity, even though there were no mistakes. For the *Monthly* prototype, participants made errors when creating an activity; many wanted to click on the specific date instead of the (+) button. Finally, there were mismatches with our terminology in our prototypes. Although we tried to mitigate this issue in our experimental tasks, the words “save” and “send” on the prototypes were confusing to users.

Intuitive. Participants still thought these prototypes were quite straightforward with “good logic”. Although the *Weekly* prototype was an innovative design, it incorporated universal signifiers, like dots to indicate an event. And, participants liked seeing people’s names associated with the invitations at the top. The *Monthly* prototype overall had an intuitive display of the calendar. Furthermore, participants recognized the mailbox icon quickly, understanding it to be an icon for pending invitations.

Discussion

Interpretation of results

Although none of our results reported to be significant, there was a strong trend suggesting the *Monthly* prototype supports our specific scheduling tasks better, especially from task 1. The qualitative analysis also revealed common mistakes that may explain the longer completion time for the *Weekly* prototype, including lack of signifiers causing the scrolling issues.

For tasks 2 and 3, the interaction effect was not significant, however, there was a trend resembling a learning effect. Although the *Monthly* prototype seemed to result in less errors for task 2, the number of errors were the same by task 3 which users were to complete immediately after task 2. *Monthly* went from 1 error to 1.5 errors on average, which may be explained by the pop-up issues for some participants. *Weekly* did not have the pop-up, and it went from 2.5 to 1.5 errors between task 2 and 3. Given the scrolling issues, we conclude that participants made more errors with *Weekly* for task 2 as they were unaware they had to scroll down to complete the rest of the steps. By task 3, participants knew they had to scroll with *Weekly*, providing a possible explanation for the decrease in number of errors. Overall, we conclude that the increase in errors for *Monthly* may be due to the pop-up and not necessarily the workflow for task 3 itself, while the decrease in *Weekly* may be due to users learning how to

complete task 2. However, the results from task 2 were only different by one error, and the t-test results were not significant, so we cannot conclude whether *Monthly* supported task 2 better than *Weekly*.

Finally, *Weekly* showed a higher level of intuitiveness for task 4, but it was not significant being only slightly higher than *Monthly* (by less than 0.5/10). Thus, we cannot conclude that *Weekly* was preferred over *Monthly* for the invitation display and accepting an invitation workflow.

Overall, we cannot determine whether one prototype supported scheduling tasks better than the other as none of our hypotheses were supported. However, we believe there is potential in determining a significant difference or a stronger trend after readjusting and repeating our experiment. Much of our results were due to potential nuisance factors, like having issues with pop-ups, lack of scrollbars, etc., which were not factors we wanted to measure. However, this gave us valuable feedback for our prototype designs overall, improving our future reiterations.

Impact for practitioners

The findings in this experiment support the importance of user-centered design. From our field study to our experiment, user feedback impacted our design approach and reiteration the most. We changed much of our scope and focus after understanding users' real needs and limitations in our field study. And, from our experiment, their feedback helped us understand more of their behaviours, abilities, and limitations, influencing where and how we should improve our prototypes for future studies.

Future Work & Recommendations

Based on our findings, the next steps would be to reiterate on our design solutions and perform the experiment again. To modify our experiment, we would recruit more participants, as a larger sample will provide a better representation of the population and thus, more accurate results. We would counterbalance the order of tasks 2 and 3 to mitigate potential learning effects from these tasks. Furthermore, we would conduct the experiment on a mobile device to increase construct and internal validity. By using a mobile device, the behaviours and findings from the experiment is a truer reflection of the real tasks. We would also include interview questions focused on aesthetic / visual preference, as this would be the next factor to consider when usability reports to be relatively the same. To reiterate our design, we recommend the following:

Weekly. To prevent scrolling issues, we recommend adding scrollbars where necessary and including half of the content to indicate there is more to see (i.e. half of the invitation icons; half of the locations display). Also, moving the "cancel" button to the top left and the "save" button to the top right would likely match users' models more, and also increase visibility of these affordances. And, we recommend designing a dropdown or thinner open text field for the activity type, as having a large open field could be too ambiguous. Although there were no errors during the experiment, we believe it may have been due to the nature of our Figma prototype since users only had to click on the open field text to fill it out, but we would need to

test this. In our updated experiment, we would want users to type in or select from a dropdown.

Monthly. Although pop-up issues were not measured in our experiment, we want to fix this issue by making the pop-up more obvious through colour, texture, and size, but also allow users to click anywhere outside the pop-up to close it. Furthermore, we recommend creating the affordance to click on the dates in a calendar to create an event and not just through the (+) button. Finally, we recommend testing designs for the invitation cards, using different colour, font styles and sizes, etc., to determine how users would best look through their invitations. Currently, we highlight the activity type by making the font larger (see figure 1b), but a participant mentioned they tend to look through the names. Thus, testing different designs and providing the appropriate interview questions can help us determine which design optimizes usability.

Conclusions

Overall, we did not find any statistical significance between the two prototypes in terms of effectiveness or intuitiveness. However, we can suggest that the *Monthly* prototype overall is more functional and practical with its calendar layout due to the strong trend we found from a faster completion time for looking through the calendar and less number of errors when editing and creating an activity.

The qualitative data gathered also offered potential explanations for user performance and behaviours on the prototypes and tasks. A frequent observation among most participants was struggling to scroll to find an already booked activity and/or scrolling through to edit an activity in *Weekly*. In the follow-up interviews, many confirmed their struggles saying “I didn’t know I could scroll”, suggesting that it may be due to the lack of signifiers and not the actual affordance to scroll that prevented them from completing the tasks faster and/or with less errors.

We also recognize that *Weekly* is more likely to suffer from limitations since participants used a desktop instead of a mobile device and scrolling on a touchscreen compared to using a mouse or touchpad are different behaviours.

Overall, our experiment gave us key insights on where and how we should improve our prototypes before testing it again to determine any potential significance. These include, but are not limited to, adding more signifiers and incorporating colours / visual elements to help with the scrolling and pop-up issues. Once we eliminate these factors, we can obtain more representative data to better evaluate effectiveness and intuitiveness of the prototypes. The visual aspect will also help us determine which invitation display is more appealing when both prototypes ultimately function at the same level of effectiveness and / or intuitiveness for the workflow. And, we believe that building our prototypes to be more horizontally developed and more autonomous makes the workflow more flexible, creating better prototypes to test for usability, effectiveness, and intuitiveness overall in the future.

References

Ferreira Silva, R. M., Mendonça, C. R., Azevedo, V. D., Raoof Memon, A., Noll, P. R. E. S., & Noll, M. (2022). Barriers to high school and university students' physical activity: A systematic review. *PLOS ONE*, *17*(4), e0265913. https://doi.org/10.1371/journal.pone.0265913

Latham, T. (2019, July 29). *The scheduling woes of adult friendship*. The Atlantic. Retrieved April 12, 2023, from https://www.theatlantic.com/family/archive/2019/07/planning-friend-hangouts-google-calendar/594586/

Appendix A.I: Supplementary Analysis

Figure 4a: Affinity diagram for *Weekly* prototype.

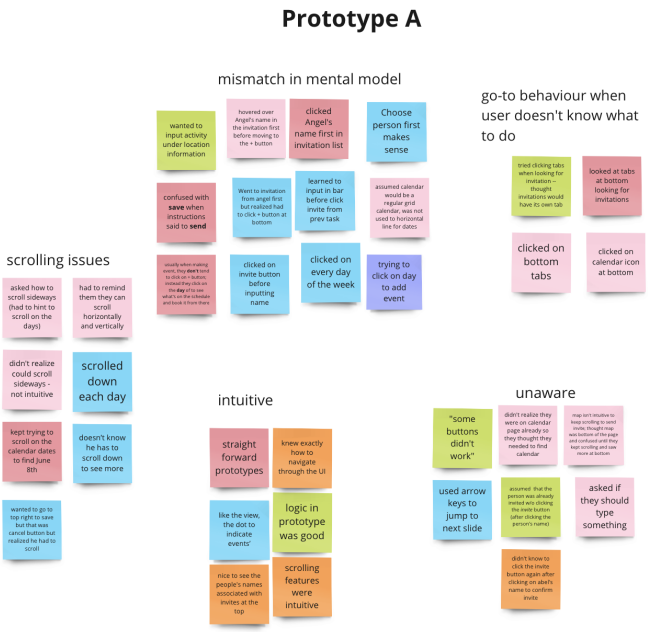
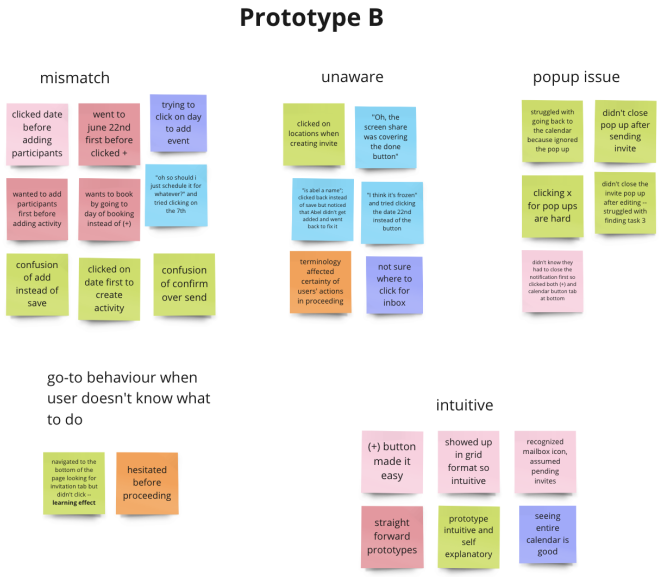


Figure 4b: Affinity diagram for *Monthly* prototype.



Appendix A.II: Experimental Materials

Scheduling Tasks

- Task 1: Fnd and click on an already booked activity in your schedule.
- Task 2: Edit the already booked activity by inviting Abel to join you and John, changing the date to June 8, changing the time to 12:30PM to 2PM, and then sending the changes.
- Task 3: Invite Angel to play Badminton with you on June 22 from 9-10AM.
- Task 4: Look through your invitations and accept the invite from Will to go play basketball on June 22 at 10-11AM and then find the event in your calendar.

Interview Questions

1. On a scale from 1 to 10, how intuitive was it to look through all of pending invites and accept one for Prototype A/B ? 1 being extremely difficult, and 10 being extremely easy.
2. How easy was it to find a booked event in the calendar for Prototype A/B from a scale of 1 to 10, with 1 being extremely difficult and 10 being extremely easy?
3. How easy was it to edit a pre-existing booked event in the calendar for Prototype A/B from a scale of 1 to 10, with 1 being extremely difficult and 10 being extremely easy?
4. How easy was it to send an invite to an existing match for Prototype A/B from a scale of 1 to 10, with 1 being extremely difficult and 10 being extremely easy?

Coding Sheet

https://docs.google.com/spreadsheets/d/1bj3A7yLSTvyx6RwMMRNKcPdSe63n9yXLPLDhG-3MqFk/edit?usp=sharing

Figma Prototypes

Weekly:
https://www.figma.com/proto/HQHuRrGnwpUic5tKlyuyxU/Find-Fit-Friends?page-id=180%3A6629&node-id=236-12326&viewport=957%2C-650%2C0.23&scaling=scale-down&starting-point-node-id=236%3A12326

Monthly:
https://www.figma.com/proto/HQHuRrGnwpUic5tKlyuyxU/Find-Fit-Friends?page-id=57%3A1303&node-id=189-10284&viewport=610%2C-91%2C0.13&scaling=min-zoom&starting-point-node-id=189%3A10284&show-proto-sidebar=1

Appendix A.III: Raw Data

Raw Data from Coding Sheet:

https://docs.google.com/spreadsheets/d/14B5rk9OmlLeL8JkkdwVowxmBoWN3KnYM69mex5KWEJw/edit?usp=sharing