

Executive Summary	3
1- Project idea: Smart TV Interface	3
The 5 W's of the problem.	3
Who	3
What	3
Where and When	4
Why	4
How	4
2- Usability fails - sampling report and project idea	4
Existing patents	5
Example Ideas	7
Usability Fails	9
User Goals	9
3- Project proposal, interface design, and user flow map	10
Mid-Fi Prototypes	10
4- Preliminary functional prototype	13
5- Guerilla Testing Results and Prototype Updates	13
Specific User Pain and Errors:	15
Appendix	16
Appendix 1, Observations for the Guerilla Testing	16
Appendix 2, Prompting users to turn on the deceive on application start-up	19
Appendix 3, High-Fidelity prototype	19
Benchmark Testing Results	22
Summary	22
Tasks and tables	22
Observations	24
Concerns	24
Solutions	25
Report on Benchmarking Data	25
Summary	25
Data Collected from Benchmark Testing	25
Data Analysis & Measures	26
Experimental design and Hypothesis	27
Summary	27
Hypothesis	27
Independent and dependent variables	27
Measuring the dependant variables	28

Anticipated confound variables	28
The appropriate statistical tests you will run for each variable	28
Experimental Protocol	28
Summary	28
Purpose	28
Questions & Tasks	28
Tasks	29
Exit Questions and User Impressions:	29
Counterbalancing	29
A/B pilot testing report	30
Summary	30
Prototype Summary & Description	30
Power Button & Remote Drawer	30
Prototype A	31
Prototype B	31
A/B Testing Results	31
Summary	31
Research Question	31
Summary of Statistics	32
Null and alternate hypothesis, and state your selected alpha	33
Independent, dependent, and confounding variables	34
Possible Type 1 and Type 2 errors in our data	34
Data type of each variable	34
T-Test results & report	34
Number of Clicks	34
Number of Errors	35
Conclusion	35
Appendices	36
Appendix 1: Prototype A for A/B Testing	36
Appendix 2: Prototype B for A/B Testing	40
Appendix 3: Final Prototype	46
Appendix 4: A/B Testing Observations	51

Executive Summary

The project is a **smartphone based remote to control your smart T.V.** In our initial user observations, we found that users had **difficulty navigating a smart T.V interface** and menus using the traditional directional pad (d-pad) navigation on a television remote. There was a large number of clicks required to navigate to various apps and through menus.

A link to our high fidelity prototype is here:

<https://xd.adobe.com/view/fd01e8fd-a611-468c-938a-004c9f152987-a852/>

1- Project idea: Smart TV Interface

This section contains the results of our initial observations about our problem, formatted as the 5 W's and How.

The 5 W's of the problem.

Who

The users that we observed were parents and siblings within our home. The age and demographic of the users varied significantly, however, the problem was found to be common amongst all the users.

What

We noticed that **users had difficulty navigating a Smart TV interface with the given controls**. The controls used were a traditional T.V remote which is ubiquitous and simple to understand. However, it is an outdated mode of input.

Problems that arose:

- Complicated navigation and impractical input (too many clicks to get to a specific application)
 - Small menu interface (requires a lot of scrolling because of the remote)
 - Typing is not practical (a lot of arrow movement on the remote and voice control is not available for all Smart TVs)

- Remote mobile applications are the same as the physical remote, which maps the same layout and carries on the same issues.

Where and When

This was found to be a problem amongst users at home but is likely to hold amongst any users of a smart T.V. It is particularly true for users who are new to a smart T.V and have not developed small “shortcuts” or muscle memory to speed the process up.

Why

As stated above, the traditional T.V remote uses a d-pad for the most control and specific key buttons such as volume control for commonly used inputs. This has not evolved to match the complex nature of traditional T.V interfaces. Televisions today are essentially computers that require complex input sequences and typing in order to be used effectively.

How

The following are potential solutions:

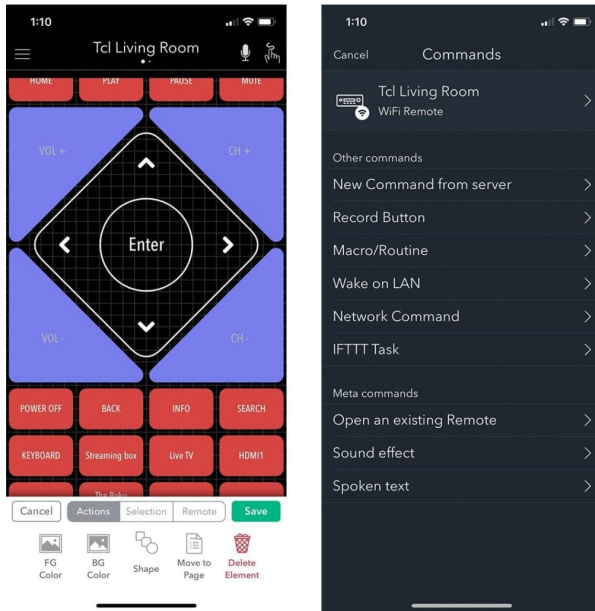
- **a mobile application that allows for more complex input**
- **a redesigned Smart TV U.I** that uses the given input (TV Remote) more effectively and intuitively
- An updated model of input in the form of **an alternative T.V Remote**

2- Usability fails - sampling report and project idea

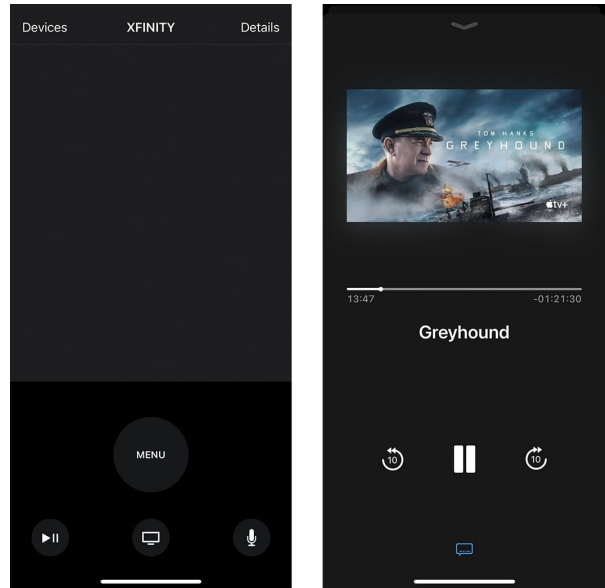
This section contains our initial research into the problem space, by looking into competitors. We then created 10 idea sketches of different aspects of the UI element, using some of them in our final design.

Existing patents

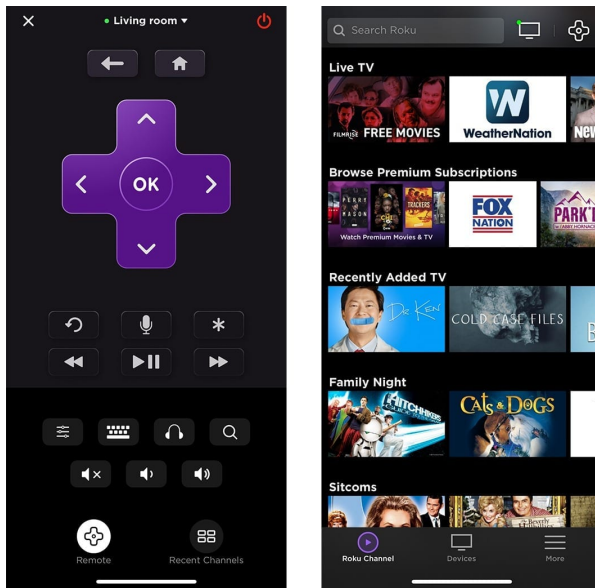
AnyMote



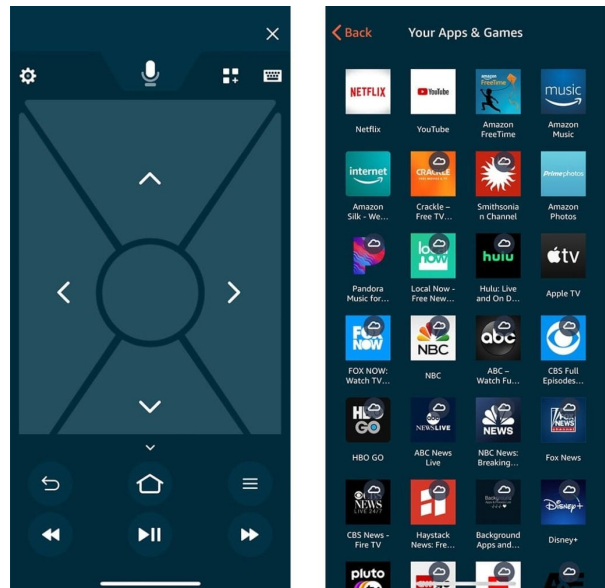
Apple TV

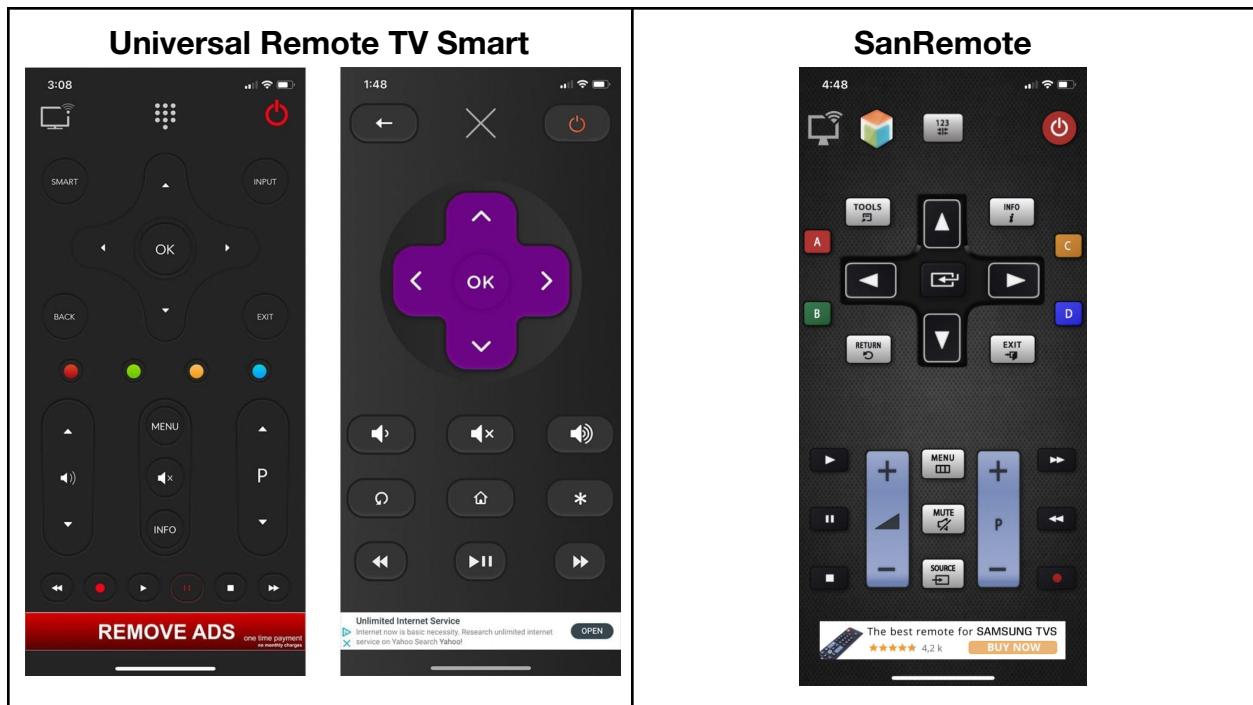


Roku TV



Amazon Fire





Reference: <https://www.hellotech.com/blog/best-tv-remote-apps-universal-streaming>

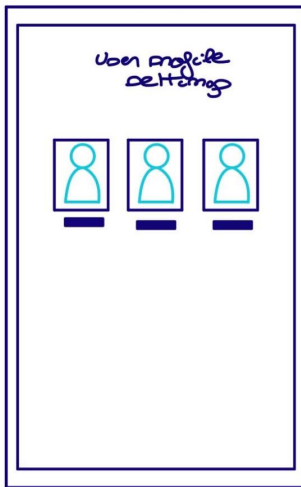
Existing applications are specific to a brand of Smart T.V, which does not allow users to use them for devices without the corresponding brand. Most of these interfaces were used as inspiration for our application, specifically adding some of the features found in these apps. However, most of these applications tackle streaming services, negating cable TV, or have a lot of inputs, without any context behind the icons.

The Anymote, Universal Remote TV Smart, and the SanRemote are all made to control multiple TVs. However, they're essentially digital versions of the traditional T.V. remote. Therefore they suffer from the same issues that the traditional T.V. remote would suffer from.

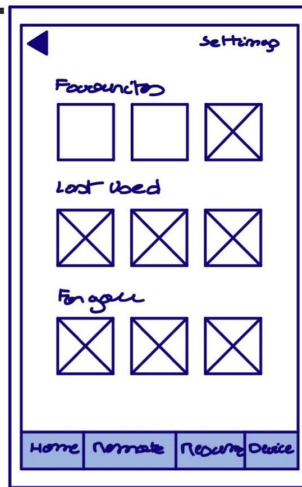
The Apple TV is extremely simplistic and has limited controls, while the Roku TV and Amazon Fire apps both have the most full range of controls. They both allow you to control the TV using a traditional style control but also manage the streaming content.

Example Ideas

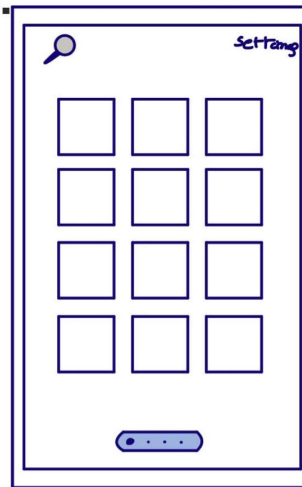
1.



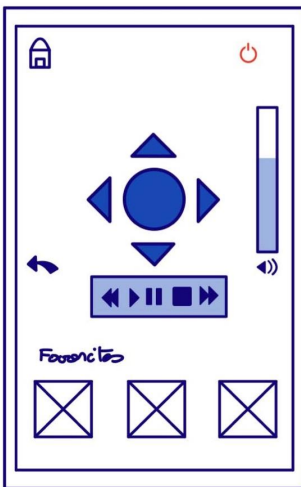
2.



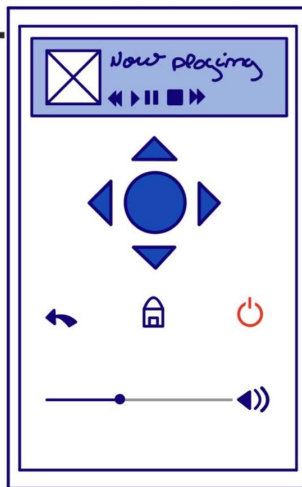
3.



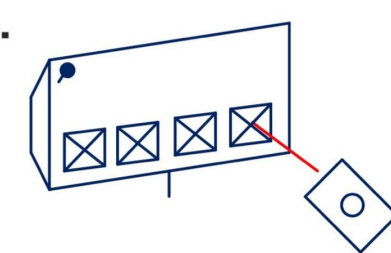
4.



5.

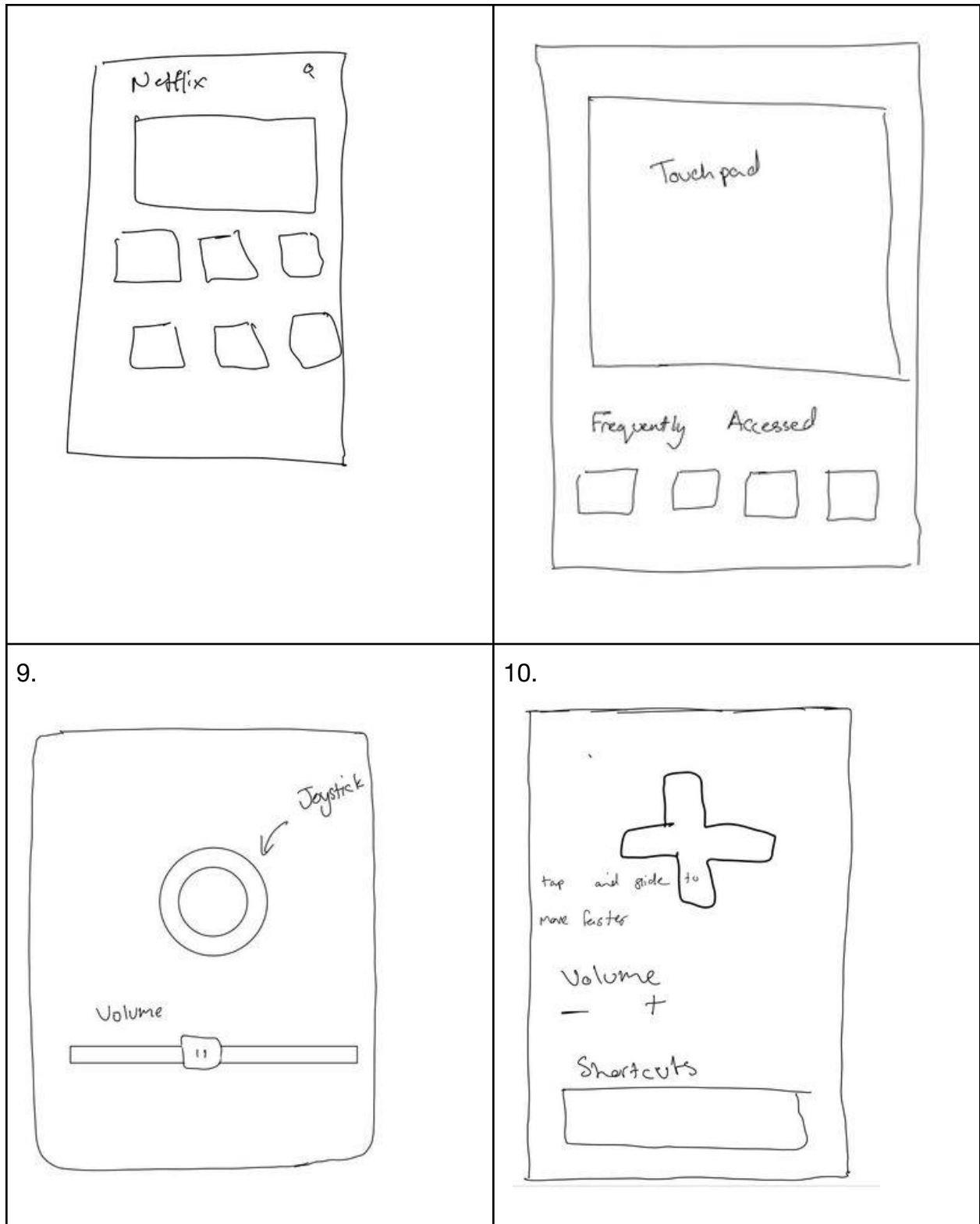


6.



7.

8.



1. User login screen
2. Potential home page

3. Full catalog of apps/channels
4. Example TV remote UI
5. Another Example TV Remote
6. Controlling a TV using motion controls
7. Example Netflix app, similar to Netflix mobile app
8. Touchpad-based controls, similar to a laptop touchpad
9. Joystick controls, including a volume slider for control
10. Alternative D-Pad control scheme, using additional types of touch interaction

Usability Fails

As discussed earlier, users have been observed struggling with Smart TV's navigation and input. These issues primarily arise with the use of the TV remote.

A traditional TV remote was designed to change the channel or volume. This made the generic up and down arrows very useful. Commonly, a number pad was included to allow for the input of a specific channel. Today, **Smart T.V's have complex settings menus and many different types of interaction that goes beyond simple up and down inputs** that were used before. This can lead to **a large number of clicks to achieve simple tasks**, such as playing a Youtube video or a movie on Netflix.

Additionally, users experienced significant difficulty **typing** on a smart T.V. The complex applications being used on a Smart T.V today require you to search through large catalogs of content by typing things out using a d-pad. The on-screen keyboard is not much help as **it takes a lot of clicks to navigate to each specific letter**. This is an inconvenient method of input. A much better option to control this would be to have a remote designed for typing.

Finally, users are prone to **commit errors while navigating a smart device with a traditional remote, specifically selecting programs**. If they were to select the wrong channel or application, they would have to go back and search for the correct one, causing a lot of unnecessary clicks. This is not optimal for the user and a digitized alternative could facilitate their navigation, thus reducing these errors.

User Goals

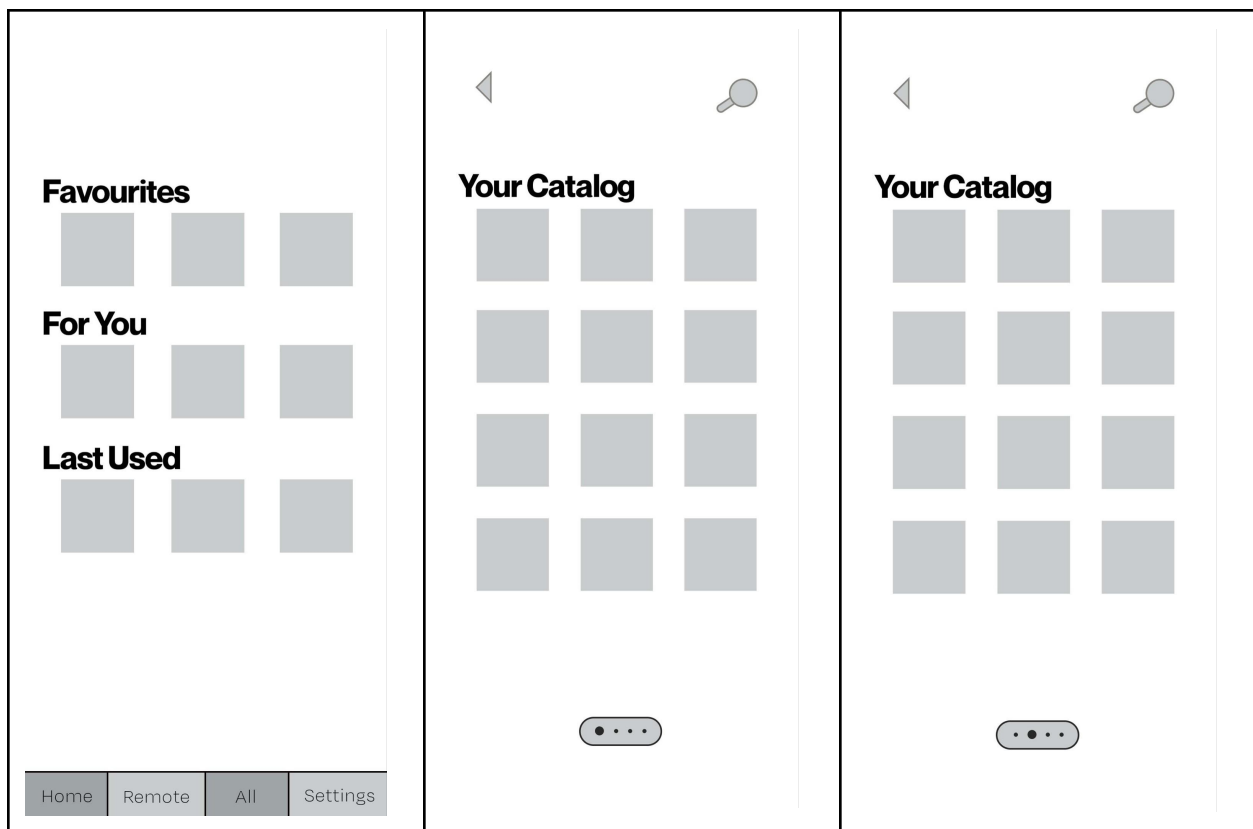
Ideally, users would navigate the application by accessing content with ease, without any complications. This should be achieved by minimizing clicks, prioritizing

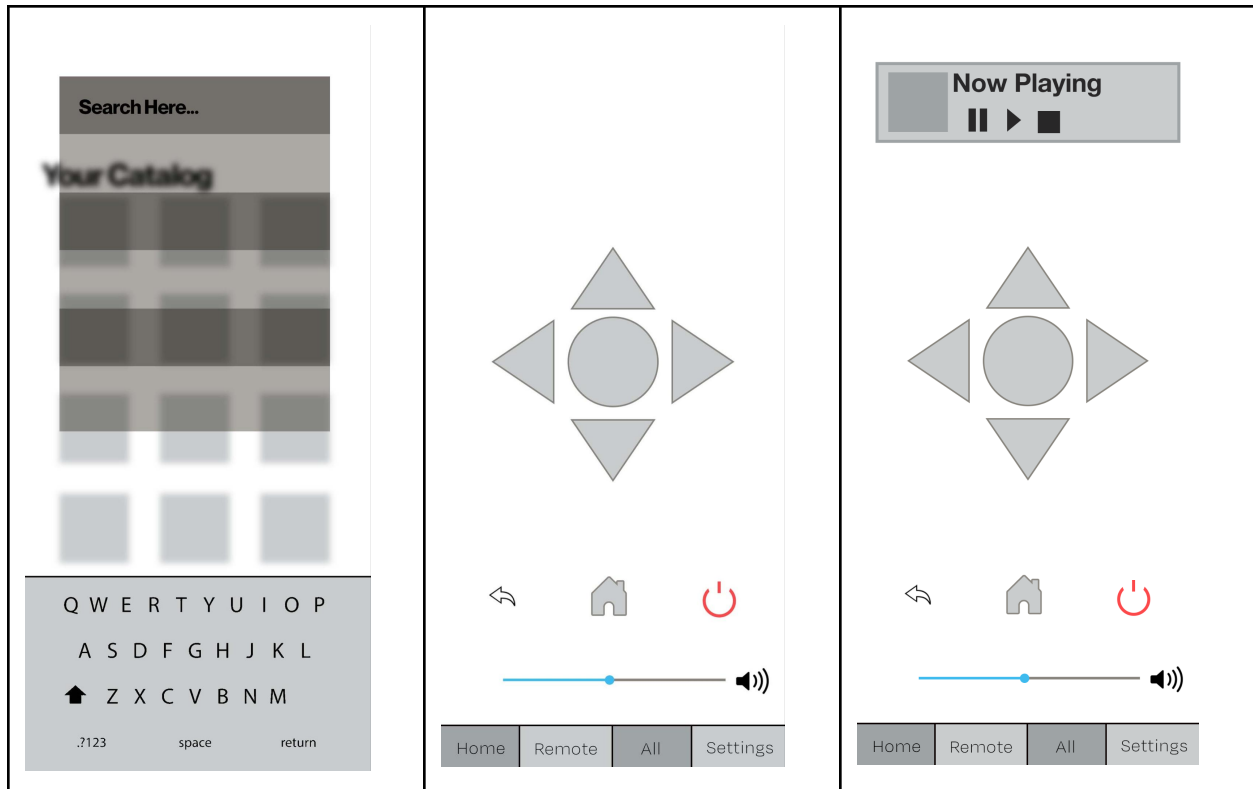
content, having an appropriate visual hierarchy and navigation cues, and facilitating users' decision-making and content discovery when navigating their device.

3- Project proposal, interface design, and user flow map

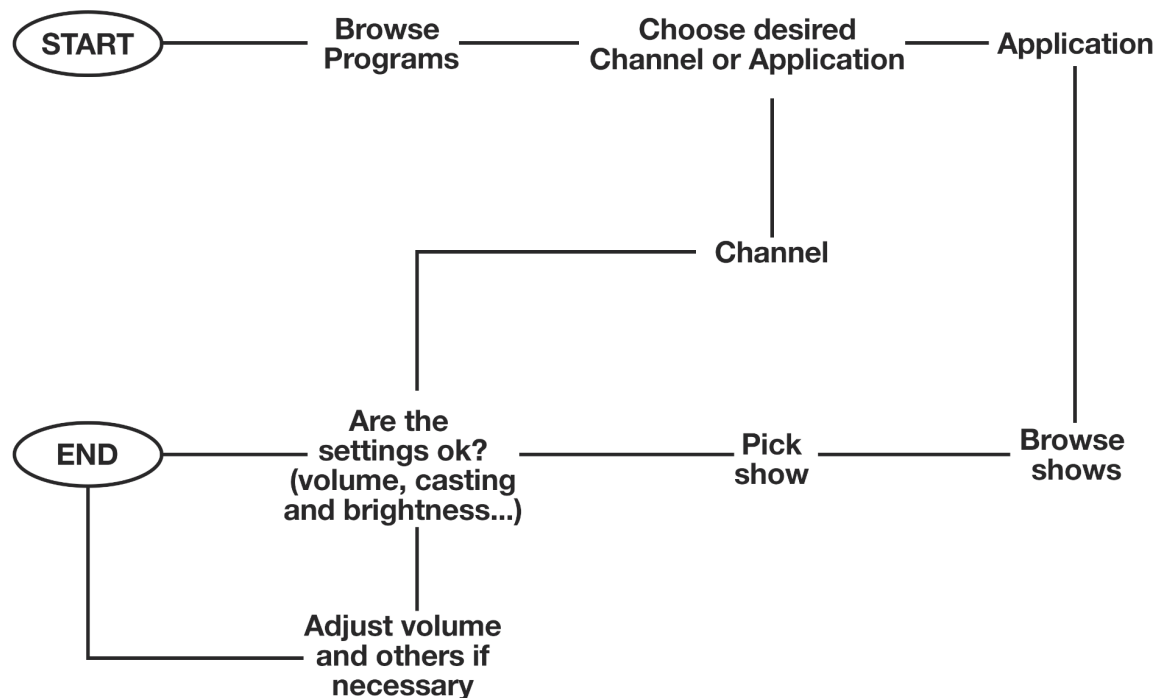
Problem: We noticed users struggling with the navigation of the interface of the Smart TV. Through our observations, users have had to rely on the arrow keys of the physical remote to navigate the screen and confirm their actions. However, this takes up more time if they do not recall the location of specific programs or applications, as the user would have to use more “clicks” than necessary. They are more prone to commit errors and the error recovery is not negligible, as it costs them time and frustrates the users. We thus seek to solve this issue by proposing an application that addresses the complicated navigations of a traditional remote due to its restrictions and limitations. We seek to mirror the functions of the remote and to make them tactile, therefore reducing user errors and relying on visual cues.

Mid-Fi Prototypes





After considering the design requirements, as well as other existing solutions within the space, this is the sample design that we have chosen to move forward with. Upon start-up, users will be prompted to turn on their devices in order to access the application. They are then redirected to the home page, which includes their ***Favorites***, ***Last Used***, and ***For You*** carousels. This creates a customized experience, as they can all be changed and adjusted, for the user, and offers additional functionality over the traditional hand-held remote. By adding customizability, users will have the option to facilitate their navigation and highlight programs and applications that they frequently use, thus reducing their clicks and search time. Moreover, the two main paths for the user are going towards the full catalog of apps under the ***All*** heading in the navigation bar, or to the pseudo-remote which is located under the ***Remote*** heading. The ***All*** section possesses all the programs and applications available on the TV. Therefore, if users need to find a specific program, they could go to the ***All*** page and either search for it by swiping through the pages or by using the search function. In addition, the ***Remote*** should encompass all the functionality found on a typical T.V remote. It seeks to mimic the traditional remote so that users could still use the features that they are familiar with.



This is a sample user flow for the application, where a user would go through our application and decide to select the desired channel or streaming application. In this flow, we are going from start to finish of the user experience. We start at **Start** and then move through **Browse Programs** to **Choose desired Channel or Application**. After giving the option to choose between the two, the flow diverges into two branches, **Application**, and **Channel**. In case they chose an **Application**, They can choose to **Browse shows** and proceed to **pick a show**. If not, they can select the desired **channel**. Both options lead the user to another choice. **“Are the settings okay?”** leads to either **“Adjust Volume and Brightness”** or directly to **End**.

This is just one sample flow of choices for the User. There are other options depending on the choices the user makes. This flow is the most universal one and will be observed throughout testing.

4- Preliminary functional prototype

This section contains an example functional prototype, essentially our first iteration of our high-fidelity prototype. We also discuss the design choices behind the functional prototype. Check Appendix 3 for screenshots of our prototype.

Prototype link:

<https://xd.adobe.com/view/fd01e8fd-a611-468c-938a-004c9f152987-a852/>

We aim to improve the interaction of users with the Smart T.V by reconsidering the traditional remote. By digitalizing the interface, errors are reduced since typing is no longer an issue to users, and selecting specific programs is easier. The first panel suggests customizability for specific users. This gives them the choice to modify the home page to their own liking, minimizing the search of common programs, as seen in the home page.

By accessing the **PROGRAMS** page, users will be allowed to access their entire catalog. If the program is nowhere to be found, they can resort to using the search tab at the top right of the page. This should not be an issue, since the keyboard on a mobile device is easier to navigate, compared to a Smart T.V.

Users can also access the **REMOTE** section, which maps the traditional remote's simplest features. This would not overwhelm users, as some of them might not be familiar with all the inputs of the traditional variant.

Finally, the **SETTINGS** section allows users to change a few visual features, such as font size and contrast. As testing progresses, more features will be added corresponding with the user's needs.

5- Guerilla Testing Results and Prototype Updates

This section focuses on our results from our Guerilla testing with our first prototype. The goal for this was to test our prototype for any clear issues with the usability and user flow. We did not have a full high-fidelity prototype completed for this, so we used a modified version of our low-fidelity drawings. These drawings were then used to make our high-fidelity prototype

(<https://xd.adobe.com/view/fd01e8fd-a611-468c-938a-004c9f152987-a852/>)

At the beginning of the testing session, we asked users for their year and program of study, as well as their familiarity with a Smart TV:

User	Year & Program of Study	Have they used a Smart TV?
1	4th year, DEM	Yes
2	4th year, CCIT	Yes
3	4th year, CS & CCIT	Yes
4	4th year, CCIT	Yes
5	4th year, CCIT	Yes

From this, we can see that our testers are all familiar with the U.I of a smart T.V. Most T.V companies have similar elements to their U.I, and almost all T.V remotes use the same style of input, containing a traditional directional pad and other buttons for a key input.

Our tasks were basic and used to test the limited functionality that we had at the time.

Tasks:

1. From the homepage, find the entire catalog
2. Locate the red channel
3. Locate the red channel through an alternate method

Testing Results:

	U1	U2	U3	U4	U5
Completed the tasks with ease					
Easily navigated the homepage					
Successfully found the ALL button					
Easily found the required information					
Was familiar with the interface					
Thought that the labels were coherently					
Expected to see more features	No	Yes	Yes	Yes	Yes

For observations, check Appendix 1

Legend

	Good - User knew what to do with relative ease
	Average - User had difficulty but accomplished the task in a reasonable amount of time
	Bad - User was unable to complete the task without administrator interference or experienced significant confusion with the tasks as explained

From our testing results, we can see that **all users experienced some difficulty with the interface** in some way. **80% of our users were expecting to see more features** in our prototype, and **an equal percentage had some difficulty with finding the required information.**

Specific User Pain and Errors:

*Confusion regarding the **ALL** button:* Certain users did not find the **ALL** button clearly due to its placement at the bottom of the page. Some had assumed that the entire catalog could be accessed through the **REMOTE** tab, however, this was incorrect. This confusion was caused by the unclear distinction between certain buttons and what differentiates them. For example, one user did not complete the first task correctly due to this confusion. The **ALL** button was not clear and was ambiguous to some, as they questioned what it was referring to and which page it would direct them to.

Lack of feedback was an issue for certain users: Certain users did not know if they had gone to the correct page due to the lack of feedback. For example, one user made an error and went to the **REMOTE** tab instead of the **ALL** for the first task. The screens and the buttons were not clear, which caused the user to be confused and click on multiple buttons as a solution.

User frustrations due to the interface: Certain pages were vague and lacked the corresponding indicators and did not display the information clearly. For example, in the case of the **REMOTE** page, most users did not know its purpose and limitations. Due to its nature, users assume that it had to be used during all three tasks. However, this was not the case, and due to the lack of feedback, users had difficulty interacting with the **REMOTE** feature.

Based on the feedback, we have decided to add a better information hierarchy in order to make the interaction more user-friendly. In addition, prompting the user to turn the Smart T.V on through the app would be helpful, and would avoid confusion, as certain users did not know whether the device was on or not (Appendix 2). Another possibility would be to direct the user directly to the remote screen, upon opening the application.

Appendix

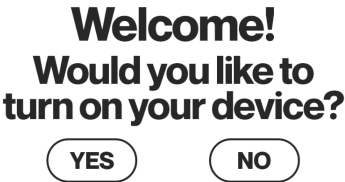
Appendix 1, Observations for the Guerilla Testing

	User Struggles	Positive Feedback
User 1	• Had a hard time with their navigation (specially the first screen) and was confused: the user was not familiar with the interface and did not complete the task successfully (until probed and with a lot of guidance) • They were unable to read and tried to change the channel using the arrow keys (from the REMOTE screens) • The REMOTE and ALL buttons were not clear to the user, even though they had used a Smart TV before • UI should be consistent and the prototype wasn't interactive, which confused the user (feedback issue)	• Ease with recovery and could figure out the navigation of the channels catalog (however, it wasn't the most intuitive experience)
User 2	• Had to clarify to the user that the interface was not the TV's • Did not find the ALL button easily, however they navigated the catalog successfully after being probed	• Navigated the interface and completed the tasks accordingly • Appreciated the remote and favorites features • Simple and easy design

	• A new feature could be the inclusion of a cast button (where the user could mirror their phone to the TV through the application)	
User 3	• Had difficulty reading the titles and the buttons in the navigation bar • Inclusion of mic options and voice commands • Could multiple phones connect to the same TV? Is the inclusion of channel numbers possible? Can the user control volume besides from the remote screen? • The REMOTE screen was vague to the user	• Completed the tasks accordingly and successfully navigated the interfaces • The home page is helpful with a lot of option for the user
User 4	• Highlight the ALL on the homepage with a visual indicator (it should be more pronounced) • Include input source changes for the TV (such as HDMI, USB, etc...) • Integrate streaming platforms and different applications • Would a top navigation bar work in this context?	• Completed the tasks accordingly and successfully • Inclusion of the favorites feature is helpful and appreciated • User appreciated the simple interface • User appreciates the bottom navigation bar (more intuitive and comfortable)
User 5	• Incorporate a SEARCH button on the homepage	• Successfully completed the tasks accordingly and with ease

	• Rename the ALL to ALL CHANNELS (could be confusing to some users) • Claimed that some icons were questionable	• Noticed the ALL button easily and clicked on it after hearing the first task • Streamlined UI (everything is where the user wanted it to be) • Appreciated the personalization (LAST USED & FOR YOU)
--	---	--

Appendix 2, Prompting users to turn on the deceive on application start-up



Appendix 3, High-Fidelity prototype

WELCOME!

Who's Watching?



Ellie



Moe



Izzy



Klara

WELCOME, Ellie



Favourites



Recently Seen



Your Streaming Apps



Recommended For You



- HOME
- PROGRAMS
- REMOTE
- SETTINGS

ALL PROGRAMS



Apple TV+



BBC One



Bloomberg TV



Cartoon Network



CBC



CBS



City TV



CNN



CP24



Crave



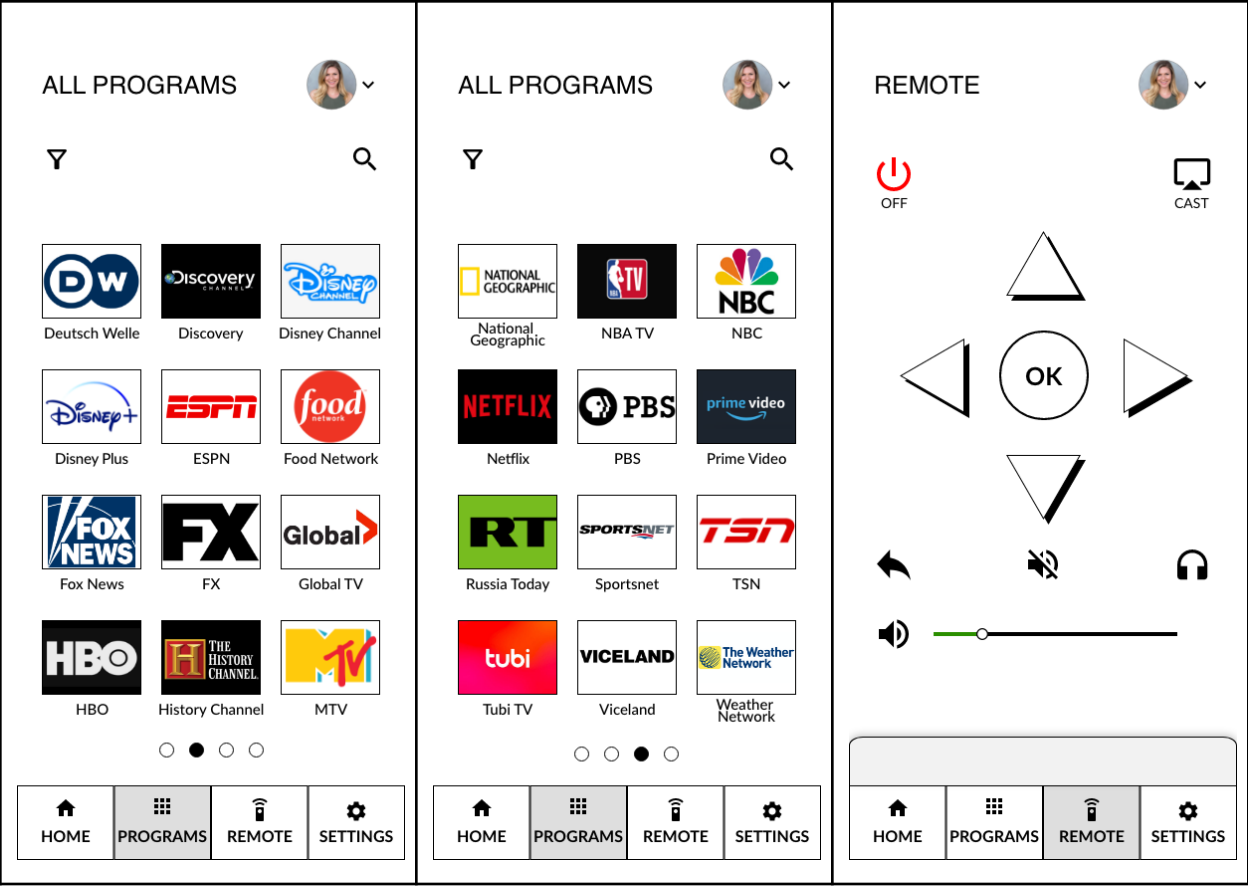
CTV

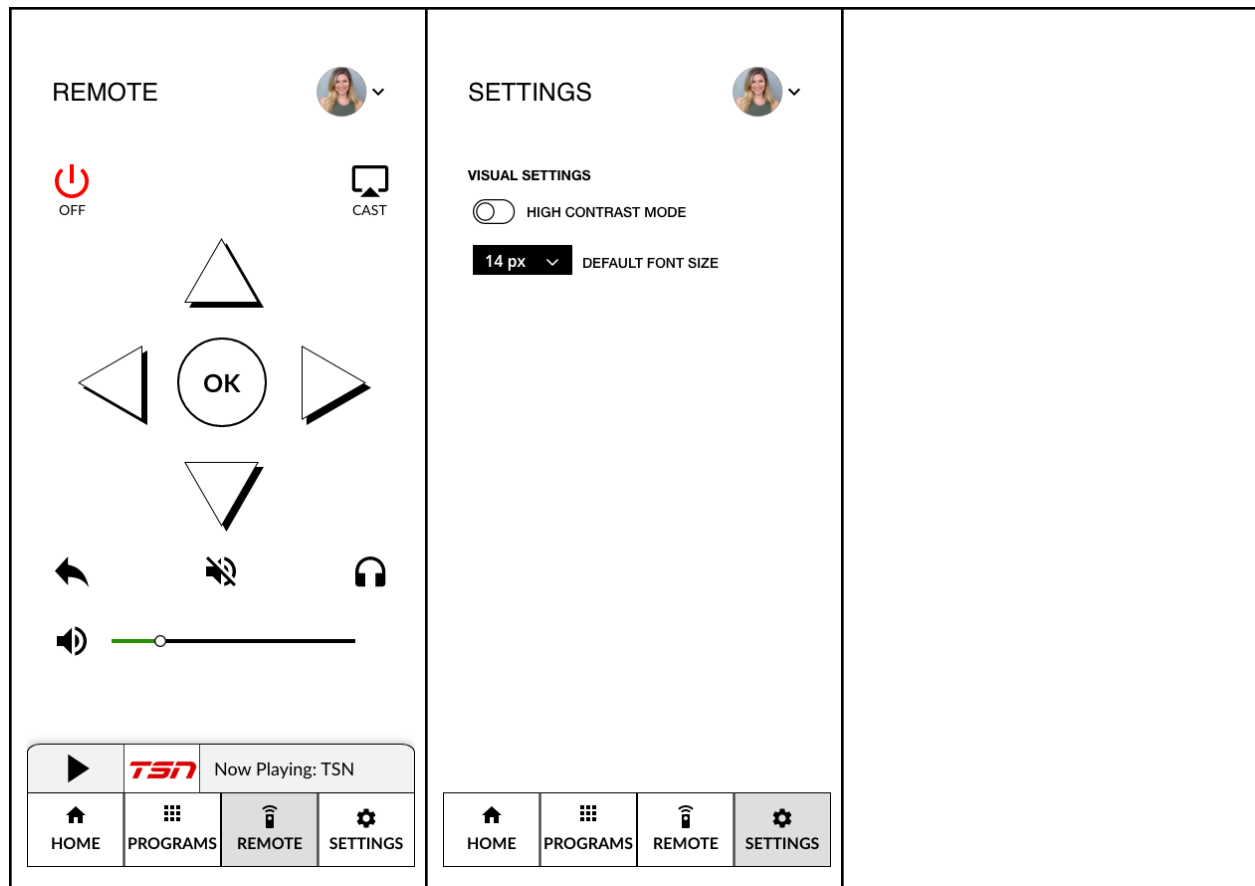


DAZN



- HOME
- PROGRAMS
- REMOTE
- SETTINGS





Benchmark Testing Results

Summary

In this report, we present the data from our benchmark testing session and address the issues that some users experienced with solutions and an improved prototype. The initial prototype was subjected to a usability testing session where we asked users to perform certain tasks.

Prototype: <https://xd.adobe.com/view/6e3274a1-c470-44b4-9c73-c0dc31637838-3b61/>

Tasks and tables

We asked users to follow the following tasks:

1. Log in and turn on the TV
2. Find the NBC channel
3. Adjust the volume
4. Change channel to Netflix

In order to **measure the deviations**, we listed expected paths:

Task 1: Login, Remote, Power On

Task 2: Programs (directs to Page 1), Page 2, Page 3, NBC

Task 3: Remote, Adjust Slider

Task 4 (*Since the User is already on the Remote page*): Return, Right Arrow to Netflix, Confirm
(For Task 4, the shortest path would be: Home, Netflix. However, we wanted to see if Users would interact with the remote to navigate the Smart TV).

We measured **two metrics**: the number of clicks and errors (or deviations from an expected path). The table below illustrates the shortest path in contrast to the path that the user took and if they committed any errors:

	Expected Path in Clicks	User 1	User 2	User 3	User 4
Task 1	3	3	7	3	5
Task 2	4	4	5	4	5
Task 3	2	2	2	2	2
Task 4	3	3	3	9	3
Total	12	12	17	18	15

Table 1, Deviations and Number of Clicks for Each User

	Home	Programs	Remote	Settings
Errors per page	6	4	0	1

Table 2, Total Errors per Page

User Number	Total Errors	Total Clicks
1	1	12
2	3	17
3	1	18
4	3	15

Table 3, Total Errors and Total Clicks per User

Observations

For Task 1, two out of the four users managed to follow the shortest path to complete the task. Between the other two users, User 2 continued to click on programs hoping that the TV would turn on automatically, and User 4 searched for the power button on the wrong pages.

For Task 2, none of the four users achieved the shortest path. User 1 and User 3 however, can be excused considering that the shortest path involves jumping from page 1 to page 3, a counterintuitive action. Logically, the users first went to page 2 to ensure that NBC was not on that page. User 2 and User 4 however, both took a similar path, which involved going to the home page first before going to the programs page. This could shed some light on user expectations about where to find channels.

For Task 3, all four of the users managed to complete the path in the least number of clicks, showing that the task was very intuitive to complete through the application

For Task 4, none of the four users achieved the shortest path. User 1, 2, and 3, however, found the second shortest path. The extra step taken showed that the users were unaware that Netflix could be accessed through the home page, and instead opted to go through the programs page. User 3, on the other hand, went through multiple pages, and crossed over programs page 3 and the home page, both of which had the Netflix icon. The user struggled to find Netflix for 9 clicks. This statistic could be evidence that even on pages where Netflix is present, it is not immediately visible to the eye and one may overlook it. This could also just be a data outlier however.

Concerns

Settings: Users noted that the location was not optimal and the button was not very visible for certain users. One user stated that the settings icon could be under PROFILE, where a drop down or a pop-up would appear with different tabs. Another user also stated that the search icon was not visible, and a labeled search bar instead of an icon would be more visible.

Complicated navigation: Users were expected to change the program (Task 4, from NBC to Netflix) with the remote function. However, all users either went back to the entire catalog or to the home page, where they found the required application.

User confusion: 3 out of 4 users struggled by clicking on the wrong page or inputs. A few struggled to turn the TV on and asked if playing a program would turn the TV on. This confusion was caused by a lack of feedback and unclear labels. Users would not know whether the TV was on (unless stated otherwise). This caused users to struggle to turn the TV

on, as they were not prompted by our application to turn their device on start-up. Some users thought that by selecting a program, the TV would start, however, they had to go to the remote and turn it on from there, leading to more steps than necessary to complete their tasks.

Solutions

To address the issue of turning on the TV, the power button will be put on all pages instead of just the remote page. A status button with constant feedback will help users and inform them whether their device is on or off.

Report on Benchmarking Data

Summary

This section contains a statistical analysis of our benchmark testing. During testing, we measured the number of clicks and compared that to an expected path. We also measured if the user deviated from the expected path.

During our benchmark testing, we measured **two metrics**: the number of clicks and errors (or deviations from an expected path). The tables below illustrate the results:

Data Collected from Benchmark Testing

	Expected Path in Clicks	User 1	User 2	User 3	User 4
Task 1	3	3	7	3	5
Task 2	4	4	5	4	5

	Expected Path in Clicks	User 1	User 2	User 3	User 4
Task 1	3	3	7	3	5
Task 3	2	2	2	2	2
Task 4	3	3	3	9	3
Total	12	12	17	18	15

Table 1, Deviations During Tasks and Number of Clicks for Each User, per Task

Legend

<i>User Deviated</i>	
<i>User Did Not Deviate</i>	

	User 1	User 2	User 3	User 4
Task 1	0	4	0	2
Task 2	0	1	0	1
Task 3	0	0	0	0
Task 4	1	1	3	2

Table 2, Total Errors of Users per Task

Data Analysis & Measures

Number of Clicks per Task									
Task 1	Task 2	Task 3	Task 4			Task 1	Task 2	Task 3	Task 4
Mean	4.5	Mean	4.5	Mean	2	Mean	4.5		
Standard E	0.957427	Standard E	0.288675	Standard E	0	Standard E	1.5		
Median	4	Median	4.5	Median	2	Median	3		
Mode	3	Mode	4	Mode	2	Mode	3		
Standard D	1.914854	Standard D	0.57735	Standard D	0	Standard D	3		
Sample Var	3.666667	Sample Var	0.333333	Sample Var	0	Sample Var	9		
Kurtosis	-1.289256	Kurtosis	-6	Kurtosis	N/A	Kurtosis	4		
Skewness	0.854563	Skewness	0	Skewness	N/A	Skewness	2		
Range	4	Range	1	Range	0	Range	6		
Minimum	3	Minimum	4	Minimum	2	Minimum	3		
Maximum	7	Maximum	5	Maximum	2	Maximum	9		
Sum	18	Sum	18	Sum	8	Sum	18		
Count	4	Count	4	Count	4	Count	4		
Number of Errors per Task									
Task 1	Task 2	Task 3	Task 4			Task 1	Task 2	Task 3	Task 4
Mean	1.5	Mean	0.5	Mean	0	Mean	1.75		
Standard E	0.957427	Standard E	0.288675	Standard E	0	Standard E	0.478714		
Median	1	Median	0.5	Median	0	Median	1.5		
Mode	0	Mode	0	Mode	0	Mode	1		
Standard D	1.914854	Standard D	0.57735	Standard D	0	Standard D	0.957427		
Sample Var	3.666667	Sample Var	0.333333	Sample Var	0	Sample Var	0.916667		
Kurtosis	-1.289256	Kurtosis	-6	Kurtosis	#DIV/0!	Kurtosis	-1.289256		
Skewness	0.854563	Skewness	0	Skewness	#DIV/0!	Skewness	0.854563		
Range	4	Range	1	Range	0	Range	2		
Minimum	0	Minimum	0	Minimum	0	Minimum	1		
Maximum	4	Maximum	1	Maximum	0	Maximum	3		
Sum	6	Sum	2	Sum	0	Sum	7		
Count	4	Count	4	Count	4	Count	4		

Measures of centrality and spread

For task 4, it can be seen in the table that the users have had the expected number of clicks but are still being considered as having deviated, and the reason for that is that they did not use the 'remote' feature of the app (which is what was expected from them).

The mean for the number of errors per task demonstrates that the participants were able to accomplish task 2 and task 3 with ease, while they struggled with task 1 and task 4. The high standard deviation and variance of task 1 and task 4 also demonstrates that while these tasks were very easy for some participants, they were considerably difficult for others - this could provide insight about the task as well as about the participants.

We have observed three insights during our session:

- 1) Some users struggled to find out how to turn the TV on - the navigation to the remote for accessing the power button was not obvious to all users
- 2) Users were mostly comfortable finding TV channels and adjusting the volume of the TV
- 3) Users struggled with finding out that the remote could be used to navigate the TV - they preferred going back to the programs page and finding their desired programs from there

Experimental design and Hypothesis

Summary

This section contains an outline of our experimental design for A/B testing. It includes our hypothesis, independent variables, as well as the plan for our statistical analysis.

Hypothesis

As mentioned in previous iterations, we seek to improve the navigation of a smart television through an application that allows the user. Therefore, we assume that our application facilitates and improves the navigation of the smart TV, compared to the traditional manual remote. We hypothesize that Prototype B reduces the amount of unnecessary clicks and errors, compared to our alternative layout, Prototype A. In this case, our null hypothesis would be that there is no difference between A and B when it comes to error and click reduction.

Independent and dependent variables

In order to conduct our experiment, we have set our independent variables to be the design changes between our two prototypes, A and B (Appendix 1 and Appendix 2). With these changes, we have set our dependent variables to be the number of clicks, number of deviations from our expected path and the users' comfort between prototypes.

Measuring the dependant variables

In order to adequately measure our variables, we will be counting the number of steps that it takes for each user to accomplish a certain task, while also pre-planning the expected path for these steps. In addition, we will take note of any deviations users take in order to determine whether or not they committed an error. Finally, we used both of these metrics to determine how comfortable the user was with our application.

Anticipated confound variables

Since our prototype has been tested in class before, we anticipate some users to be familiar with the interface. Therefore, we will need to counterbalance our task

order in order to avoid that, and screen our applicants in order to determine whether or not they already tested the application before.

The appropriate statistical tests you will run for each variable

T-tests will be run on each variable to determine the probability of the results achieved, allowing for an objective measure of the application's effectiveness

Experimental Protocol

Summary

This section contains our entire experimental protocol for A/B testing. It includes a description of our tasks, how we plan to counter-balance, as well as the questions we want to ask before and after our testing.

Purpose

As suggested in previous iterations, we noticed users struggling with the navigation of the interface of the Smart TV. We therefore seek to reduce the amount of unnecessary “clicks” and facilitate error recovery for users.

Questions & Tasks

Before the testing, we plan to ask users if they have tested our prototype before (“Have you tested with our prototype before?”), whether it was an older or newer prototype. With this question, we could determine whether or not their experience was affected by a past experience.

Tasks

We outlined the following tasks for our users in order to measure specific metrics, such as number of clicks, number of errors and comfortability with the application:

1. **Power ON the T.V**
2. Task A: You would like to change to/turn on the *NBC channel*
3. Task B: Someone just walked in, you need to mute the TV. Now that they have left the room, you want to unmute it.
4. Task C: You would like to change to/turn on *Disney +*
5. **Power OFF the TV**

After the completion of the tasks, we plan to ask our user what they liked about the prototype that they interacted with or if they had any issues with it.

Exit Questions and User Impressions:

- Post-Screening Questions:
 - After seeing both, which features did you prefer?
 - Which prototype do you like more? Why?
 - What did you not like from the other one?

Counterbalancing

To avoid learning effect, we divided users into 2 batches and counterbalanced the tasks as it follows:

Prototype A then B	Task	Task	Task
User 1	A	B	C
User 3	C	A	B
User 5	B	C	A

Prototype B then A	Task	Task	Task
User 4	A	B	C
User 2	C	A	B
User 6	B	C	A

If we have more users, we will go back to User 1,

A/B pilot testing report

Summary

This section contains a report on our first round of A/B testing, including issues that we faced and adjustments that we made.

Initially, our A/B testing went well. Our protocol was holding up. We made a mistake in our initial counterbalancing approach, and had to swap the protocol order for Users 2 and 4, but that was the only big mistake we made.

The largest issue that we had was a lack of qualitative information being relayed to us by the participants. We did not have as much info from each participant as we needed. As we continued through the testing, we noticed that additional screening questions would benefit us, as well as asking questions throughout the testing, primarily when swapping between prototypes. The largest adjustment that we made was including the question “What did you like or dislike about this prototype?” in between the swap to each prototype. We also added a few additional questions to the end of our testing such as “What did you like for each prototype?”

The biggest difference these adjustments made is additional qualitative data, as we will be asking more questions. We will also get a better idea of how the differences between the prototypes affects the participants’ impression of each. Overall, besides the additional questions, everything else about our testing went well.

Prototype Summary & Description

Power Button & Remote Drawer

For both prototypes (Appendix 1 & 2), we kept the layout of the homescreen from our older prototypes and added a power button to the top right corner for signaling. We have also noticed in previous testing sessions that the remote feature was not used, even when users were probed to. Thus, we remove the standalone remote tab from our nav bar, and replace it with a drawer, which users could access from anywhere on the application. Instead of it solely being a digital mapping of the manual remote, the remote drawer includes the most important features from the remote in a single interface. Through this, we anticipate users to solely use the remote when necessary, reducing the number of clicks, and avoiding redirection to a different page.

Prototype A

For this prototype (Appendix 1), the home screen displays the different programs and channels under customizable categories in a horizontal row layout, separating each category with a title. This allows users to easily locate what they have placed in their respective categories, allowing them to perform actions with less clicks.

In addition, the programs page is a grid layout that displays all of the programs into separate pages. To go through the pages the user would ideally swipe, however, with our current prototype, they would have to click to the next page.

Prototype B

The main difference between A and B (Appendix 3) is the homepage. In our B prototype, the program categories are listed under categorized hexagons. The sole indicator is the label, which indicates what the user could potentially find inside of it. This was done to reduce clutter from our A prototype, while keeping the information accessible and categorized. However, it could potentially come to the cost of more clicks and errors.

The program page holds the same information hierarchy as A, but instead of a swiping motion, the user would have to scroll in order to go through the channels.

A/B Testing Results

Summary

This section contains the results and analysis from our A/B testing. We present a statistical analysis, as well as our final conclusions made, and changes made to the prototypes based on these results.

Research Question

As stated in previous iterations of our project, we seek to reduce the amount of unnecessary “clicks” and facilitate error recovery for users navigating their smart television with their traditional remote. We therefore suggest an application that facilitates the navigation of their device in order to achieve their goals through less clicks and with comfortable error recovery. In order to achieve this goal and to determine the features that will be used in our final prototype, we tested 10 users with two different prototypes: A and B.

Summary of Statistics

Users answers to our post-screening survey							
Users	Tested before	Question 1	Question 2	Question 3	Question 4	Question 5	Question 6
1a	Yes	5	4	5	5	4	5
2b	No	5	5	3	3	5	5
3a	No	4	5	4	4	5	5
4b	Yes	5	4	5	5	4	5
5a	Yes	4	3	4	5	4	4
6b	No	4	5	3	3	5	4
7a (1a)	No	3	4	2	2	4	4
8b (2b)	No	4	3	3	4	4	4
9a (5a)	Yes	5	3	5	5	3	4
10b (4b)	Yes	5	5	4	5	4	5
Number of clicks during each task							
Users	Task Aa	Task Ba	Task Ca	Task Ab	Task Bb	Task Cb	
1a	4	4	2	3	4	1	
2b	3	3	1	3	3	5	
3a	3	6	1	3	3	4	
4b	3	3	3	3	8	5	
5a	4	3	3	4	3	2	
6b	6	3	2	3	3	4	
7a (1a)	6	6	2	3	3	4	
8b (2b)	4	3	1	3	8	2	
9a (5a)	5	7	2	3	3	2	
10b (4b)	4	2	3	3	5	2	

User answers to the survey (top), Number of clicks during each task (bottom)

Number of Errors	User 1	User 2	User 3	User 4	User 5	User 6	User 7	User 8	User 9	User 10
Task Aa	0	0	0	0	0	4	2	0	4	4
Task Ba	0	0	0	3	0	0	4	0	4	0
Task Ca	0	0	0	0	0	1	1	1	1	0
Task Ab	0	0	0	0	1	0	5	0	0	0
Task Bb	0	0	0	5	0	0	0	6	0	2
Task Cb	0	3	2	3	0	2	3	0	0	0

Number of errors per user for each task

During our A/B testing sessions, we took notes of the number of clicks for each task (per user) and had our users fill a post-testing survey with the following questions:

- **Question 1:** How satisfied were you with your overall interaction?
- **Question 2:** Compared to other alternatives, how efficient was your interaction with the "rectangles" (Prototype A)?
- **Question 3:** Compared to other alternatives, how efficient was your interaction with the "hexagons" (Prototype B)?
- **Question 4:** How easy was it to find the corresponding channels (through the catalog) on the "hexagon" prototype?
- **Question 5:** How easy was it to find the corresponding channels (through the catalog) on the "rectangular" prototype?
- **Question 6:** How would you rate your interaction with the remote drawer?

In the survey, users had to complete a 5-point likert scale describing their interaction with the

questions listed above. We then collected the following results:

Task Aa		Task Ba		Task Ca		Task Ab		Task Bb		Task Cb	
Mean	4.2	Mean	4	Mean	2	Mean	3.1	Mean	4.3	Mean	3.1
Standard Err	0.35901099	Standard Err	0.53748385	Standard Err	0.25819889	Standard Err	0.1	Standard Err	0.65064071	Standard Err	0.45825757
Median	4	Median	3	Median	2	Median	3	Median	3	Median	3
Mode	4	Mode	3	Mode	2	Mode	3	Mode	3	Mode	2
Standard Dev	1.13529242	Standard Dev	1.69967317	Standard Dev	0.81649658	Standard Dev	0.31622777	Standard Dev	2.05750658	Standard Dev	1.44913767
Sample Vari	1.28888889	Sample Vari	2.88888889	Sample Vari	0.66666667	Sample Vari	0.1	Sample Vari	4.23333333	Sample Vari	2.1
Kurtosis	-0.7089774	Kurtosis	-0.8349535	Kurtosis	-1.3928571	Kurtosis	10	Kurtosis	0.43736659	Kurtosis	-1.6996005
Skewness	0.66062275	Skewness	0.84857943	Skewness	0	Skewness	3.16227766	Skewness	1.41023841	Skewness	0.06024382
Range	3	Range	5	Range	2	Range	1	Range	5	Range	4
Minimum	3	Minimum	2	Minimum	1	Minimum	3	Minimum	3	Minimum	1
Maximum	6	Maximum	7	Maximum	3	Maximum	4	Maximum	8	Maximum	5
Sum	42	Sum	40	Sum	20	Sum	31	Sum	43	Sum	31
Count	10	Count	10	Count	10	Count	10	Count	10	Count	10
Question 1		Question 2		Question 3		Question 4		Question 5		Question 6	
Mean	4.4	Mean	4.1	Mean	3.8	Mean	4.1	Mean	4.2	Mean	4.5
Standard Err	0.22110832	Standard Err	0.27688746	Standard Err	0.32659863	Standard Err	0.34801022	Standard Err	0.2	Standard Err	0.16666667
Median	4.5	Median	4	Median	4	Median	4.5	Median	4	Median	4.5
Mode	5	Mode	5	Mode	5	Mode	5	Mode	4	Mode	5
Standard Dev	0.6992059	Standard Dev	0.87559504	Standard Dev	1.03279556	Standard Dev	1.10050493	Standard Dev	0.63245553	Standard Dev	0.52704628
Sample Vari	0.48888889	Sample Vari	0.76666667	Sample Vari	1.06666667	Sample Vari	1.21111111	Sample Vari	0.4	Sample Vari	0.27777778
Kurtosis	-0.1461039	Kurtosis	-1.7337294	Kurtosis	-0.8956473	Kurtosis	-0.5216011	Kurtosis	0.17857143	Kurtosis	-2.5714286
Skewness	-0.7801058	Skewness	-0.2234505	Skewness	-0.2723191	Skewness	-0.8628233	Skewness	-0.1317616	Skewness	0
Range	2	Range	2	Range	3	Range	3	Range	2	Range	1
Minimum	3	Minimum	3	Minimum	2	Minimum	2	Minimum	3	Minimum	4
Maximum	5	Maximum	5	Maximum	5	Maximum	5	Maximum	5	Maximum	5
Sum	44	Sum	41	Sum	38	Sum	41	Sum	42	Sum	45
Count	10	Count	10	Count	10	Count	10	Count	10	Count	10

Measures of spread and centrality, for the post-survey results (bottom) and the number of clicks (top)

For Task C (a and b), users noted that it was a learning curve. If they were not familiar with it during their first test, they successfully performed the task correctly during the second session. Moreover, most of our users verbally stated that the remote was not very obvious at first, but found the tasks easy to achieve after learning how to interact with it. However, we do realize that it lacked signaling. Therefore, we suggest adding a bar on the drawer (Appendix 1) to signal users that they can swipe it, or click on it, in order to access these features.

In addition, upon being asked (Question 3 & 5), most users preferred the programs page from our B prototype, due to the ease of the infinite scrolling feature.

Null and alternate hypothesis, and state your selected alpha

Our null hypothesis claims that there is no difference between our Prototypes A and B when it comes to the amount of clicks and amount of errors, with our alternate hypothesis claiming that Prototype A affords less clicks and errors than Prototype B. In order to reject our null hypothesis, we selected an α of 0.05, which we will be using during our T-Test.

Independent, dependent, and confounding variables

Throughout our testing, we have set our independent variables to be defined by the design changes between our two prototypes, A and B (Appendix 1 and Appendix 2), which sets our dependent variables to be the number of clicks, number of deviations from our expected path and the users' comfort between prototypes. However, the users' familiarity with our application could impact their testing, making it easier for them to complete certain tasks, therefore impacting the results.

Possible Type 1 and Type 2 errors in our data

A possible type 1 error could be a false positive result, which states that there is an increase in the number of clicks and errors for A, compared to B. A possible Type 2 error could refer to a false negative result, claiming that there is no difference in the number of clicks and errors between A and B.

Data type of each variable

T-Test results & report

We have chosen a t-test to compare the means of two groups, specified at the beginning of each report summary.

Number of Clicks

After running a T-Test on *Task Aa* and *Task Ab*, we got a p-value of **0.0174**. With an α of 0.05, we can determine whether or not to reject our hypothesis. Since our p-value is smaller than our α , we can reject our null hypothesis, which claims that there is no difference between both prototypes. Therefore, Prototype A affords less clicks and errors than Prototype B.

We ran another T-Test for Tasks *Ba* and *Bb* in order to see if our null hypothesis was correct when it came to our *Programs* page. Our p-value was around **0.0841** for this data set, which exceeded our α of 0.005. In this case, we failed to reject our null hypothesis, which entails that there is no difference between both prototypes when it comes to the number of clicks when it comes to the navigation of the programs page.

Number of Errors

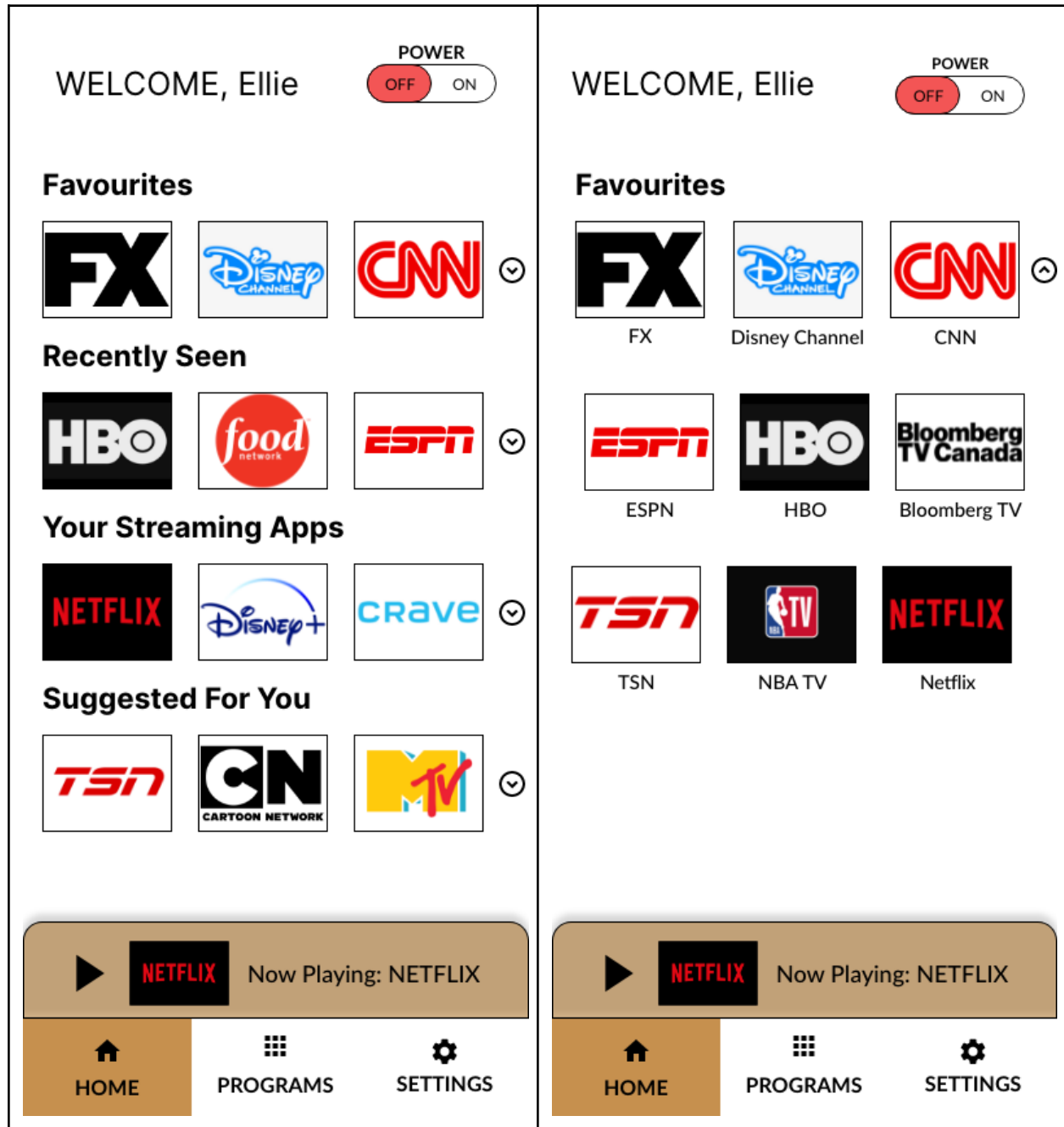
For both of our groups, (Task Aa and Task Ab, Task Ba and Task Bb), the p-values are equal to **0.3182** and **0.0946** respectively. Both of our p-values exceed the alpha of α , which rejects our null hypothesis that there is no difference when it comes to the number of errors between both prototypes.

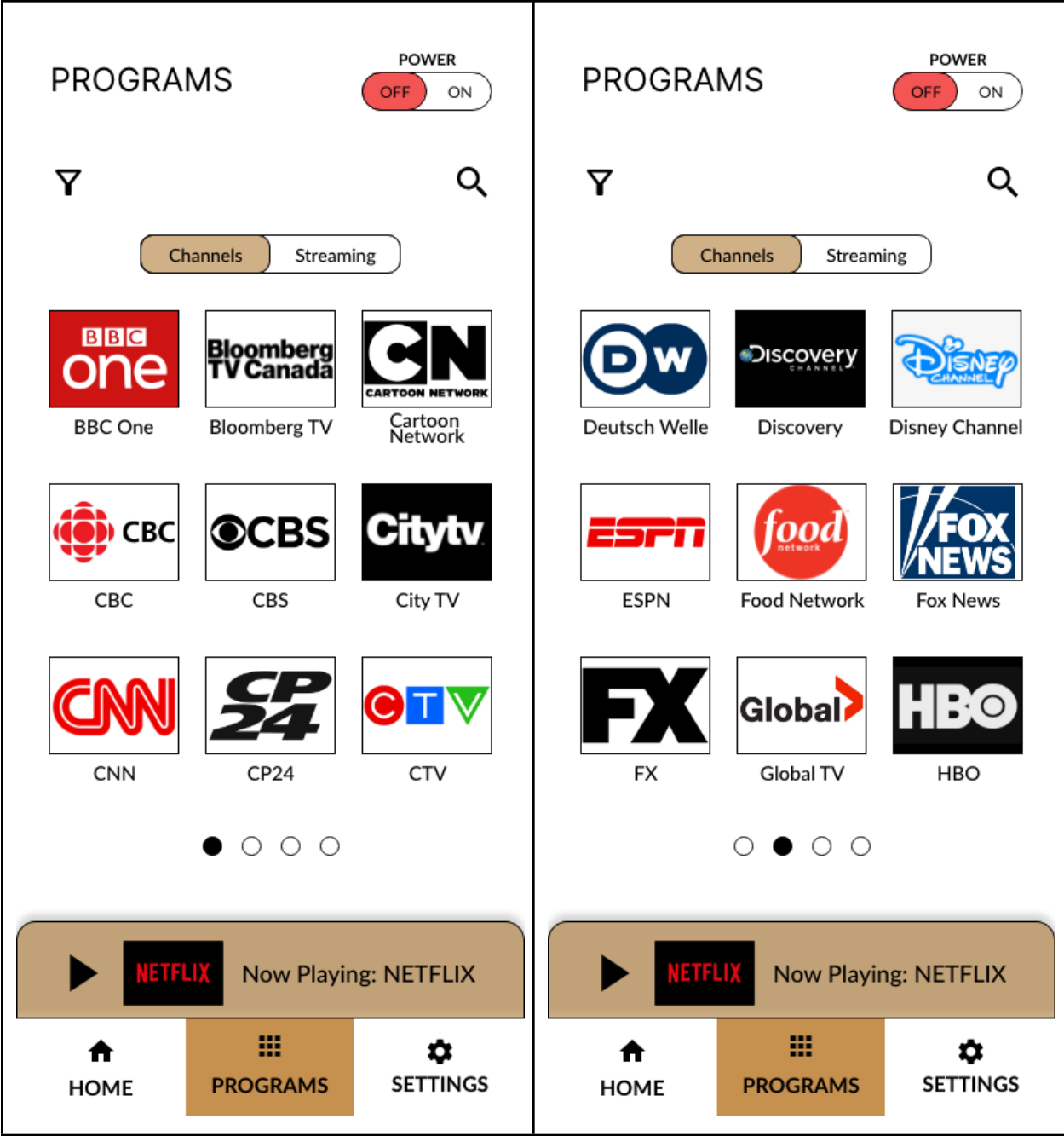
Conclusion

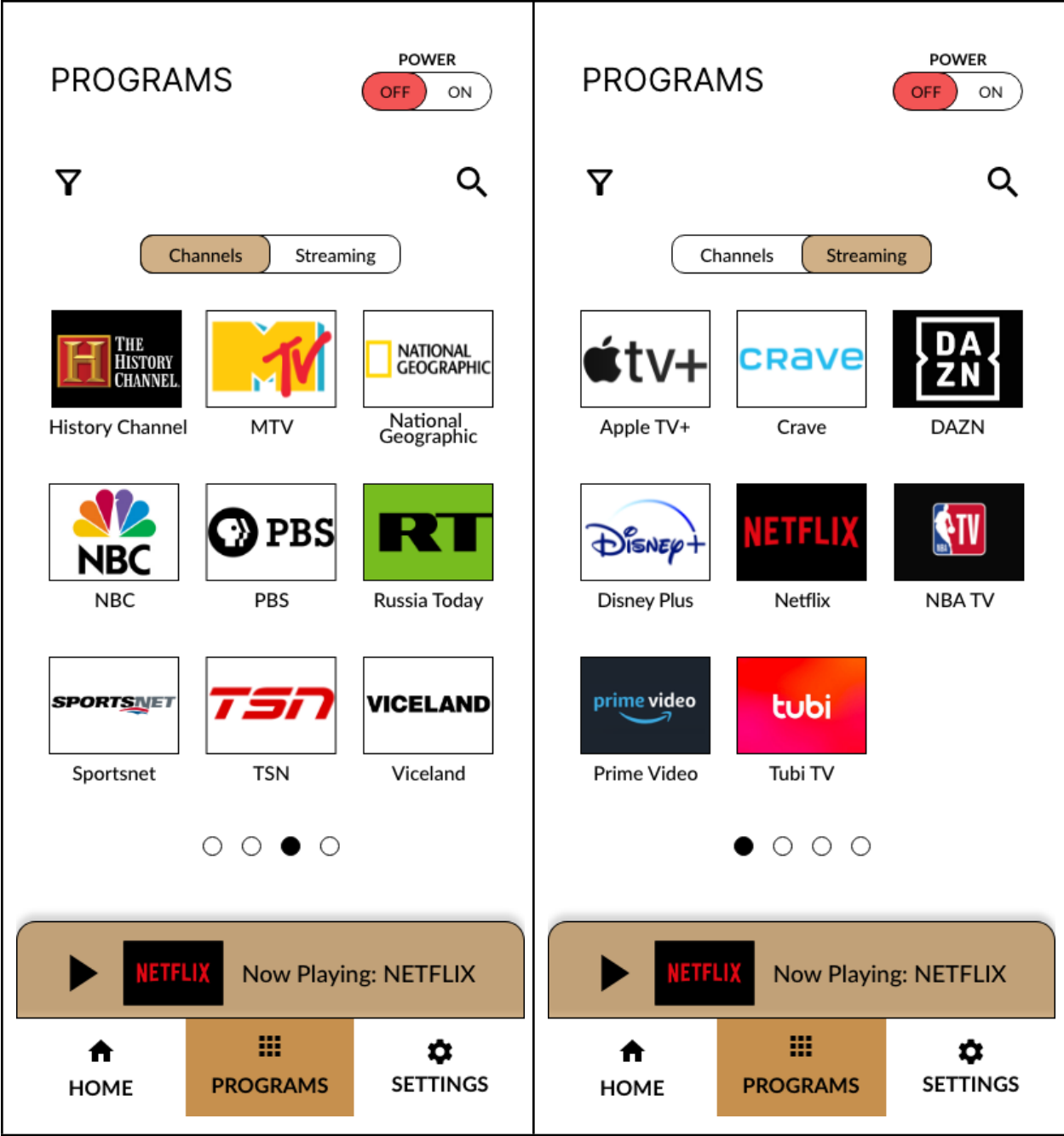
We can thus conclude that the home page of A afforded less clicks and errors, due to the visible information that was missing from the hexagons in prototype B. In addition, whilst our tests conclude that there is no difference between both program pages, our users affirmed through the questionnaire that they preferred the interface of B, due to the infinite scrolling offering more comfort. Since there was no difference between both when it came to error reduction, we have solely decided to focus on the features that promise on reducing the total amount of clicks for our final prototype.

Appendices

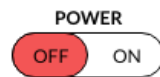
Appendix 1: Prototype A for A/B Testing







SETTINGS



VISUAL SETTINGS

☐ HIGH CONTRAST MODE

14 px ▾ DEFAULT FONT SIZE

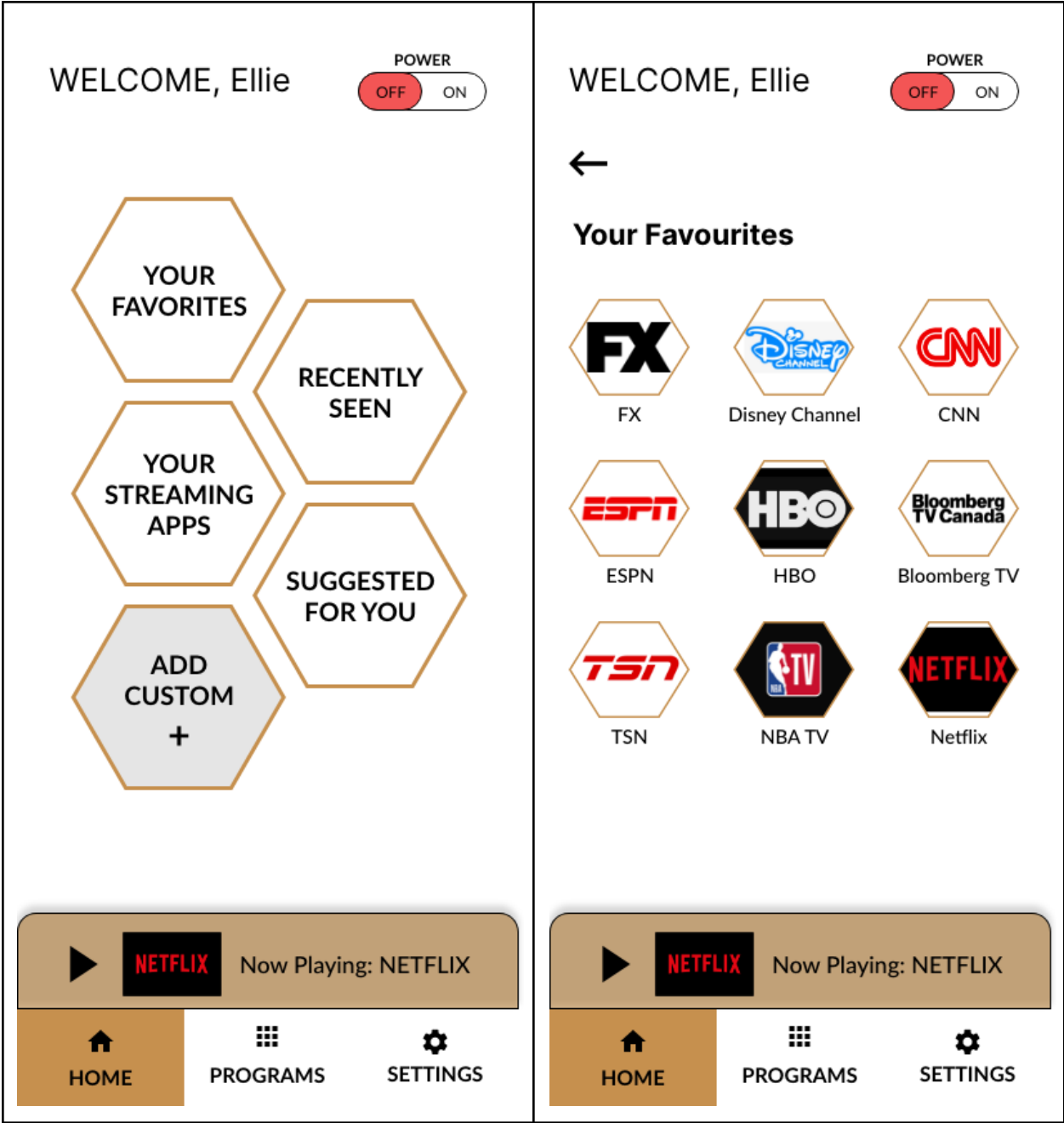
  Now Playing: NETFLIX

 HOME

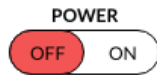
 PROGRAMS

 **SETTINGS**

Appendix 2: Prototype B for A/B Testing



WELCOME, Ellie



Recently Seen



HBO



Food Network



ESPN



MTV



Global TV



National
Geographic



The Weather
Network



Duetsch Welle



CP24



NETFLIX

Now Playing: NETFLIX



HOME

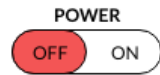


PROGRAMS



SETTINGS

WELCOME, Ellie



Your Steaming Apps



Apple TV+



Crave



DAYZ



Prime Video



Viceland



Disney+



Tubi TV



Netflix



NETFLIX

Now Playing: NETFLIX



HOME



PROGRAMS



SETTINGS

WELCOME, Ellie



Suggested For You



SportsNet



Russia Today



PBS



The History Channel



Fox News



Discovery Channel



CBC



Citytv



NETFLIX

Now Playing: NETFLIX



HOME



PROGRAMS



SETTINGS

PROGRAMS



Channels

Streaming



Apple TV+



Crave



DAZN



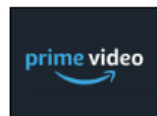
Disney Plus



Netflix



NBA TV



Prime Video



Tubi TV



NETFLIX

Now Playing: NETFLIX



HOME

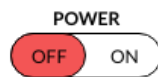


PROGRAMS



SETTINGS

SETTINGS



VISUAL SETTINGS



 Now Playing: NETFLIX

 HOME

 PROGRAMS

 SETTINGS

PROGRAMS

POWER
OFF ON



Channels Streaming



BBC One



Bloomberg TV



Cartoon Network



CBC



CBS



City TV



CNN



CP24



CTV



Deutsch Welle



Discovery



Disney Channel



ESPN



Food Network



Fox News



FX



Global TV



HBO



History Channel



MTV



National Geographic



NBC



PBS



Russia Today



Sportsnet



TSN



Viceland



NETFLIX

Now Playing: NETFLIX



HOME

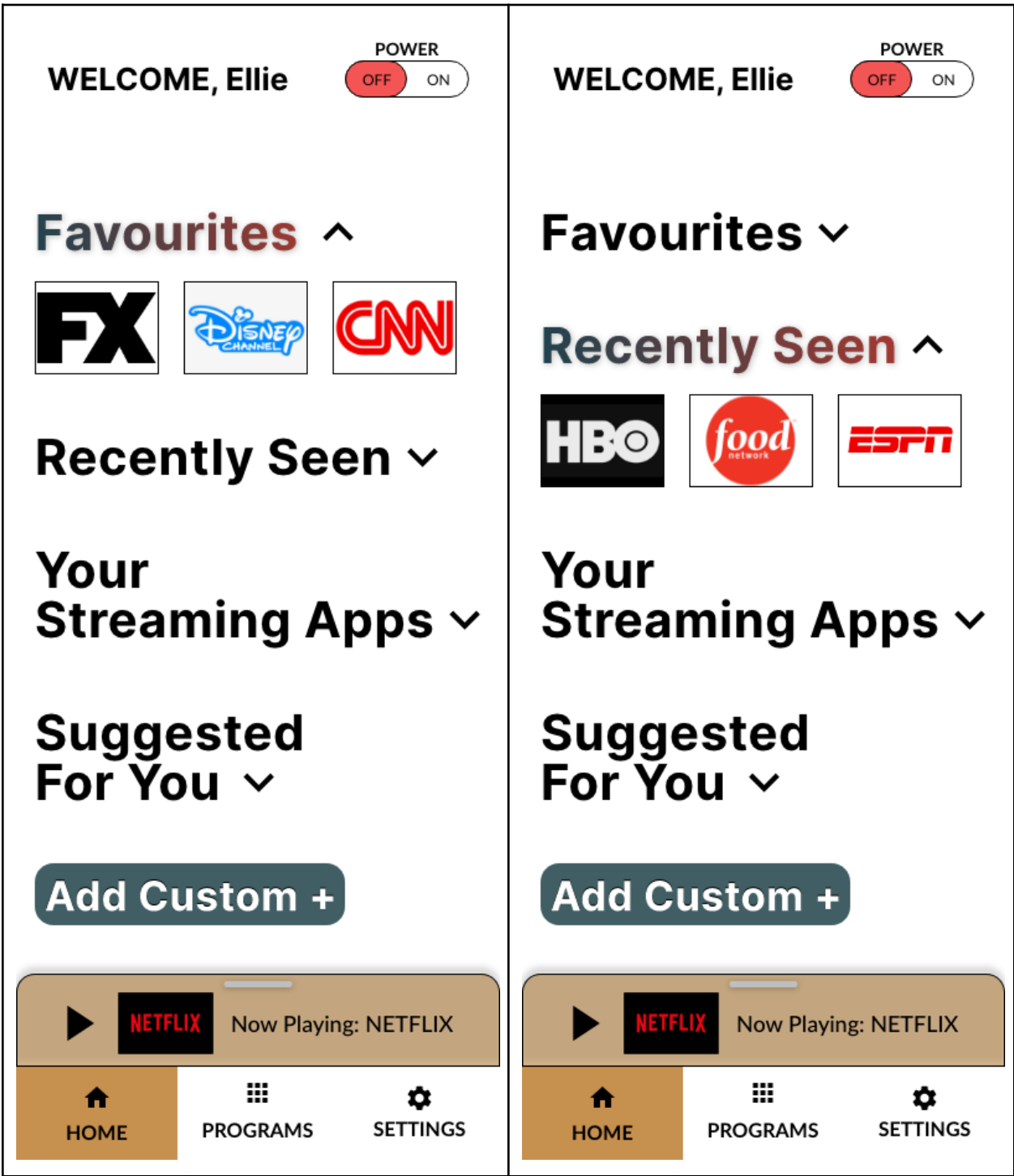


PROGRAMS



SETTINGS

Appendix 3: Final Prototype



WELCOME, Ellie



Favourites ▾

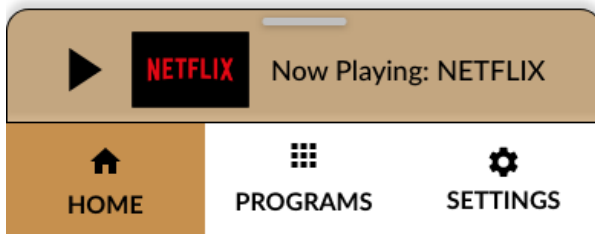
Recently Seen ▾

Your
Streaming Apps ^



Suggested
For You ▾

Add Custom +



WELCOME, Ellie



Favourites ▾

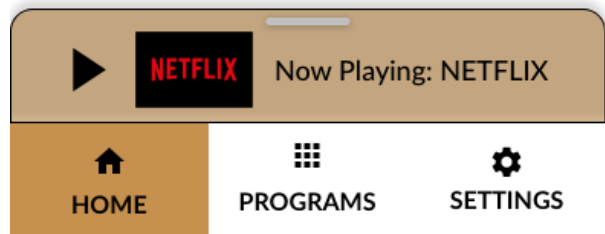
Recently Seen ▾

Your
Streaming Apps ▾

Suggested
For You ^



Add Custom +



WELCOME, Ellie



ADD NEW CAROUSEL

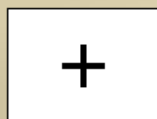
Name:

European News

Add Programs:



Russia Today



Add New

SAVE



NETFLIX

Now Playing: NETFLIX



HOME



PROGRAMS



SETTINGS

SETTINGS



VISUAL SETTINGS



HIGH CONTRAST MODE

14 px



DEFAULT FONT SIZE



NETFLIX

Now Playing: NETFLIX



HOME



PROGRAMS



SETTINGS

PROGRAMS

POWER

OFF

ON



Channels

Streaming



Apple TV+



Crave



DAZN



Disney Plus



Netflix



NBA TV



Prime Video



Tubi TV



NETFLIX

Now Playing: NETFLIX



HOME



PROGRAMS



SETTINGS

PROGRAMS

POWER

OFF ON



Channels

Streaming



BBC One



Bloomberg TV



Cartoon Network



CBC



CBS



City TV



CNN



CP24



CTV



Deutsch Welle



Discovery



Disney Channel



ESPN



Food Network



Fox News



FX



Global TV



HBO



History Channel



MTV



National Geographic



NBC



PBS



Russia Today



Sportsnet



TSN



Viceland



NETFLIX

Now Playing: NETFLIX



HOME



PROGRAMS



SETTINGS

Appendix 4: A/B Testing Observations

User	Notes	Have they tested our prototype before?	Preferred prototype
1a	A: They turned the T.V on successfully, and successfully performed all the tasks correctly. Missed the drawer at first but quickly recognized it. B: Same as A, performed everything accordingly. Stated that they preferred the homepage from A and the infinite scroll from B.	Yes	N/A
2b	Preferred the home page of A (rectangles displayed the channels/programs) due to its signaling of what was available. Preferred the catalog of B (scrolling instead of swiping is more intuitive). Missed the drawer at first (for B)	No	N/A
3a	Performed a lot of misclicks: for instance, they went to settings to mute the television. They also said that the drawer was not obvious and suggested making the remote bigger. Scrolled to the bottom of channels.	No	B
4b	B: Muted the television after heavy prompting and guiding. They also did not notice the switches for Channel/Streaming in the catalog. Suggested to add a mute button next to the drawer tab or on other key stops (for instance the top of the home page), in case other users want to mute without looking, just for their convenience. A: Expected a carousel for the home page (for the Favorites, My Streaming apps etc..), and tried to swipe. Stated that the Hexagons were counterproductive, had too much empty space and lacked information.	Yes	A

5a	A: Performed all the actions correctly and knew how to mute and unmute correctly. Stated that the interface was familiar and intuitive. B: Prefers the <i>infinite</i> scroll and stated that it was more intuitive. Preferred B because A had a lot of information on the homepage and unnecessary channels that they would not click on.	Yes	B
6b	B: Said that they were not overwhelmed with the information and everything was easy to find. At first, when they were looking for NBC, they checked the homepage since they did not notice the Channel/Streaming filter in Programs. A: Contrary to B, they clicked on Disney + from home rather than the programs page. They were more familiar with this interface and stated that they preferred the homepage of A and the program page of B. Overall, they preferred A due to its familiarity, as B did not display the information unless the user clicked on the hexagons. They also stated that they were familiar with the collapsible remote.	No	A
7a (1a)	A: At first, they read all the options and then checked favorites to find the corresponding channel. They did not know about the remote drawer and had to be probed: they expected it to ONLY “control” Netflix. This confusion caused them to go to Settings to attempt to mute the T.V. B: They clicked on all the Hexagons to figure out where NBC was and did it again to find Disney +, since they forgot where it was initially. They did not go to streaming directly because they mixed them up. They stated that A made more sense than B, but still had an overload of information. Recall V.S Recognition: the remote drawer and the hexagons. Had to investigate what the features offered (hexagons) or did not notice them at all in the beginning (remote).	No	A

8b (2b)	B: To find Disney +, they clicked on the hexagon and to find NBC, they clicked on Programs. They did not recognize the drawer at first and had to be shown that you could tap on the drawer to expand it and mute the television. A: Performed all the tasks smoothly and stated that they preferred A to B. However, a symbol should be added to the categories to indicate that they are expandable.	No	A
9a (5a)	A: Checked settings to mute the television and failed the task (Task B): they had to be prompted and then had to be shown where the remote is. The rest of the tasks were performed with ease. For the remote, they stated that <i>"it was not obvious that you could tap it"</i> and that an indicator, like a button or an icon, would help. B: They checked the hexagons for Disney + and they went to programs for NBC. They preferred B due to its homepage and the obvious labels, and they stated that A had an information overload.	Yes	B
10b (4b)	B: Performed every task with ease, besides Task B. To find the remote, they initially went to settings but then found it after being probed. They recognized that <i>Disney +</i> was a streaming app and selected it from the homepage, under <i>Your Streaming Apps</i>. A: Tried to swipe on the Programs page and then went to the Streaming tab to find <i>Disney +</i>. They appreciated features from both prototypes, such as the homepage on A and the infinite scroll from B. Pictures or visual indicators would have helped on B.	Yes	A