EXPLORING BEST PRACTICES IN USABILITY TESTING

WITH A FOCUS EYE TRACKING

by

Edie Terrell

Texas A&M University

Approved by Research Advisor: Anatol Bologan

May 2021

Major: Visualization VIST 441

TABLE OF CONTENTS

ABSTRACT

Exploring Usability Testing Best Practices and the Value Add of Eye Tracking

Edie Terrell

Department of Visualization

Texas A&M University

Research Advisor: Dr. Anatol Bologan

Department of Visualization

Texas A&M University

"Usability testing is increasingly used to evaluate and improve the design of clinical software applications"(). Usability testing can be conducted in many different fashions, all with the common goal of testing the success of an application. With the increase in usability testing comes a need to gain a better understanding of our end user. Understanding, predicting, and changing human behavior to understand what drives user decisions and creates user satisfaction. Usability testing should push further than testing the success of a prototype. There are many ways to conduct usability testing, one of which being eye tracking. By testing the way a person views and interacts with an application we are able to interpret data to better understand our user's satisfaction levels and more accurately and effectively improve design. For these reasons' usability testing offers a chance for a competitive advantage in the corporate setting. This paper aims to explore usability testing best practices and the value-add of behavioral and eye-movement tracking software and technologies and the value add they have to the corporate world. This research explores different methodologies for more effective usability testing results.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Anatol Bologan, for pushing me in design, research, and emerging technologies while also teaching me where to place my priorities.

KEY WORDS

1) UX: User Experience
2) UI: User Interface
3) Tobii: The company that creates the equipment used for eye and behavioral tracking in this evaluation.
4) iMotion: The software used to track the data collected and run the usability tests.
5) Design Thinking Process: A process followed by UI/UX designers that stresses user centric design in an iterative fashion.
6) Human Interaction: How humans interact with an application or product.

7) Usability Testing: Testing the success of a product on a user base to find changes that need to be made.
8) Usability Testing Best Practices
9) Eye-Tracking: A method used to track eye movements and fixation points.
10) Proposed Usability Test Process (PUTP)
11) Subsequential Eye-Tracking Questionnaire (SETQ)

SECTION I

RESEARCH QUESTIONS/MOTIVATIONS/ARTIFACTS

Emphasis on following the design thinking process within the UI/UX sphere has pushed usability testing to become an industry standard in recent years, following the ideals of creating an iterative process within design. The justification for usability testing is keeping designs user centric to promote higher rates of success. Every design change or decision made needs to be based on user feedback and observation. Usability testing looks at the success of a prototype or product overall, if a majority of users successfully make it through the prototype then it will have a high success rate. What about the few that weren't successful though? How do we make user testing more individualized, so we might realize the inefficiencies in our design on a more personal scale and reflect that in our research. Taking into account the differences between people and how they interact with technology is important. This paper aims to explore usability testing best practices and the value-add of eye-tracking software and technologies. This research explores different methodologies for more effective, individualistic, and inclusive usability testing results.

SECTION II

EXISTING USABILITY TESTING METHODOLOGY,

TECHNIQUES AND BEST PRACTICES

We use usability testing "to evaluate and improve the design of clinical software applications"(). Existing usability tests include but are not limited to problem solving or

formative usability testing, heuristic evaluation, benchmark testing, competitive testing, eye-tracking studies, and learnability studies.

Problem solving/ formative usability testing:

"Used during the early stages of the design and development process, the formative method identifies the issues with user interface (UI) design and provides solutions to solve those issues during the primary stages of the development process. Considered to be an excellent tool to figure out which design features are useful and which are not, formative user testing heavily influences the design decisions you will make about your product."

Heuristic evaluation:

"Heuristic evaluation is an informal method of usability analysis where a number of evaluators are presented with an interface design and asked to comment on it."

Benchmark testing:

"UX benchmarking refers to evaluating a product or service's user experience by using metrics to gauge its relative performance against a meaningful standard."

Competitive testing:

"Competitive usability evaluations are a method to determine how your site performs in relation to your competitors' sites. The comparison can be holistic, ranking sites by some overall site-usability metrics, or it can be more focused, comparing features, content, or design elements across sites.

Evaluations can take the form of expert reviews, where an experienced usability practitioner reviews the designs based on her expertise and knowledge of usability, or competitive usability testing, where users complete a set of tasks using 2 or more competing sites.

Eye Tracking studies:

Eye tracking tracks “the users pupils and their position on a screen are tracked and thus provide detailed data about the users visual attention on user interface elements. It can be used as a valuable source of information about users behaviour.”

Learnability studies:

“Learnability is one of the five quality components of usability (the others being efficiency, memorability, errors, and satisfaction). Testing learnability is especially valuable for complex applications and systems that users access frequently, though knowing how quickly users can acclimate to your interface is valuable for even objectively simple systems.”

These tests can be conducted in one or all of 3 different fashions, moderated in person, moderated remote, and unmoderated remote.

Moderated in Person:

When you are present with the user taking notes or observing in person.

Moderated Remote:

When you are not with the user in person, but are present taking notes or observing the user. This is commonly done over the computer through a system or Zoom, Skype, etc.

Unmoderated Remote:

When the user is performing the tasks of the usability test unmoderated. You are not present and the user is alone.

Usability tests do not measure simply the success or failure of a prototype or aspects of a prototype. When c resting a user test it is important to keep in mind that you will be analyzing

more than one measure. This is known as ‘triangulation’. “For example, a poorly constructed icon bar will generate:

Errors- especially picking the wrong icon on the toolbar

Slow task times- during which participants hesitate over each icon and frequently click through them looking for the one they want

Statements of frustration- participants express their feelings about not being able to learn how the icons are organized or be able to guess what an icon will do from the tool tip name

The need for assistance from the moderator

Much of the data analysis involves building a case for a usability problem by combining several measures”(235). Something interesting to note about this process is that “much of this analysis is dependent on the think aloud protocol. We depend on what participants say to help us solve the problem”(235).

Best practices that are overarching circulate around going into a usability test with a plan and direction. Being prepared and having the capacity for in the moment decisions on how to adapt questions, knowing when to dive deeper into an issue, and deviate from the script is an important practiced skill. User testing can bring to light things we might not have predicted when preparing, so being quick on your feet is favorable, lets call it improvising in a sense.

SECTION III

EYE TRACKING IS THE NEXT STEP

IN USABILITY TESTING

Understanding eye tracking and it's evolution you can look at its history. It can first be noted in 1879 when "Louis Emile Javal noticed that people do not read smoothly across a page, but rather pause on some words while moving quickly through others" (). Edmund Huey is noted to have been the first to create an eye tracking device to track eye movement in reading, publishing his findings *The Psychology and Pedagogy of Reading*. During the 1900s various improvements in eye tracking technology were made to create less intrusive observation of how we read. Eye tracking was first seen to be commercially used in the 1980s by marketing groups interested in "measuring the effectiveness of ads", "however, it wasn't really until the late 80's and early 90's that eye-tracking began to distinguish important differences in print and screen design" ().

Eye tracking looks further than a normal usability test is capable. With eye tracking we are able to uncover deeper levels of human behavior and user experience that we fail to observe in the average usability test. Eye tracking introduces another perspective to usability testing within the ability to collect data on eye movement and formulate a connection to the respective cognitive activities of the user that can be confirmed through user self reported data (). This was observed in a study done by Lynne Cook at the University or N Texas, where she found "that fixation duration and fixation frequencies are interdependent measures. Analysis of the self-reported data indicates that although users' responses are limited in what they reveal about cognitive processes, the responses generally confirm the eye movement measures" (). Eye tracking allows for us to find the optimum user experience. Within the space of usability testing

“eye tracking has been used to gain insight into human behavior that may not be available through observation or think-aloud protocols” ().

Since 2001 the leader in eye tracking tech development has been a company called Tobii Technology. Tobii’s aim is to “use eye tracking to deepen your understanding of human behavior and create new frontiers in fields such as psychology and neuroscience, infant and child development, clinical research, and more.” While also allowing “you to see things from the perspective of consumers. Whether you’re examining product placement, packaging design, advertising, or user experience, eye tracking accurately reveals what grabs attention, what influences purchase behavior, and how consumers engage with your product.” Some products to highlight from Tobii are the Tobii Pro Glasses 2, Tobii Pro Spectrum, and the Tobii X2-60 Eye Tracker.

Tobii Pro Glasses 2

The Tobii Pro Glasses 2 are notably accessible. They were “designed to enable easy, precise and efficient collection of eye tracking data in a wide variety of research scenarios.” The Tobii Pro Glasses are an extremely unobtrusive technology for data collection as seen in Fig. 1 they resemble a set of glasses referred to as a head unit that rest lightly on the face connecting to a recording device through an HDMI cable. The recording unit, as seen in Fig. 2 is“ connected to the head unit via an HDMI cable,” it “holds the battery and stores the recorded data on an SD memory card. The recording unit is controlled from a tab- let or computer running controller software.”

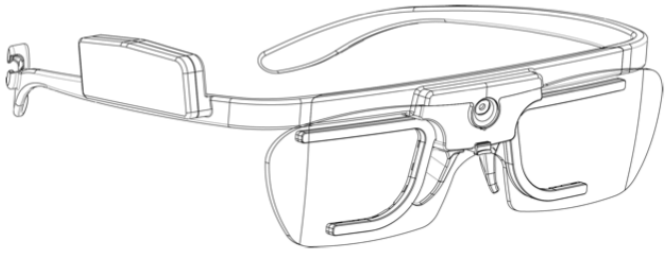

Fig. 1 Tobii Pro Glasses 2 - Head Unit

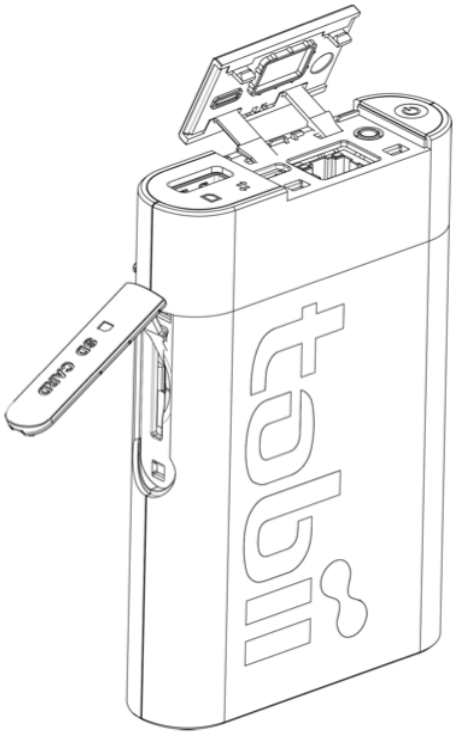

Fig. 2 Tobii Pro Glasses Recording Unit

Tobii Pro Spectrum

The Tobii Pro Spectrum is a “Screen-based eye tracker capturing gaze data at speeds up to 1200 Hz. This high-performance research system provides superior data quality and is designed for extensive research into behavior and eye movements – from fixation-based studies to micro-saccades.” The Tobii Pro Spectrum is Tobii’s most advanced eye tracking platform.

Fig. 3 The Tobii Pro Spectrum Screen

Tobii X2-60 Eye Tracker

“The Tobii X2 Eye Tracker is an unobtrusive eye tracker for detailed research of natural behavior. Its large freedom of head move- ment allows the participant to move during recording while maintaining accuracy and precision. The Tobii X2 Eye Trackers offer maximum flexibility with numerous software and stimuli setup options.”

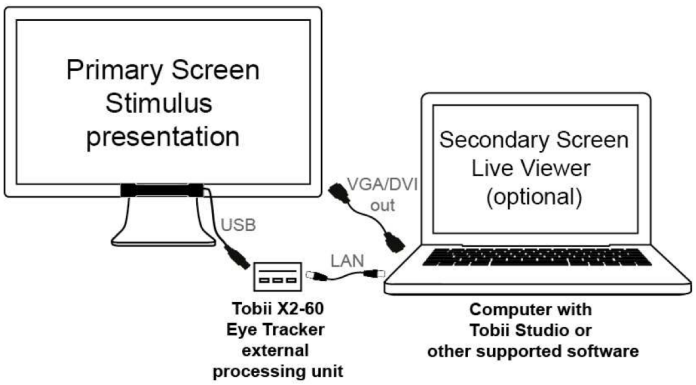

Fig. 4 Displays the set up of the Tobii X2-60 Eye Tracker

Taking into account the user's comfort and how that might affect the results of the test is important. There was a study conducted by Barker and Biers in 1994 in which they varied the presence of testing equipment in a room with the user. “They found that the presence of the equipment did not affect the participants’ performance or rating of usability of the product” (238).

SECTION IV

AT&T v. VERIZON CASE STUDY

To prove the use of user eye movement and behavioral usability test tracking I conducted a study focusing on the usability of AT&T and Verizon’s websites. The purpose of the study was to observe the usability of each site and from the data collected discern which site users prefer and if the usability of the site affected their choice of which service to sign up for. I hypothesize that if usability is a contributing factor toward if a user signs on for a service from one of these 2

service providers, then there is a value add and competitive advantage to using usability testing within corporations.

I chose to test AT&T and Verizon's sites because they are both in the B2B space offering direct competition of services. There aren't many differentiating factors as far as services offered between the two companies within the range of coverage, network quality, and devices offered. These are all within a similar disparity. It is important to acknowledge and premises that there are many factors far exceeding usability of applications and sites, as to why consumers sign up for their services.

How much of the user experience within these two businesses websites is connotated to gaining or losing customers? Since these two companies are very competitive and network coverage to a range of devices is not so much attributed to the main competitive quality of the service providers, what brings customers in? Marketing, advertising, and user experience all collectively feeding into the company reputation. User experience can range from customer support to the actual interaction between a customer and employee in a store. User experience on the companies websites isn't necessarily the dominant factor of influence to subscribe to one of these service providers, as there are other points of interaction and influence, but is still a factor to gaining or losing customers. Interaction with a website and it's usability can be the convincing moment for a user to purchase a service. A company might gain customers solely because of ease of use.

Usability testing allows us to find the pitfalls in our competitors and ourselves and directly indicates points that need improvement for the usability of our end user. Many successful leading businesses utilize usability testing such as Apple, Amazon, or Microsoft.

Products from these businesses remain competitive because users enjoy using them. It isn't hard to complete an action, ease of use is a priority. But many, specifically large corporations, do not take advantage of this. Usability testing is ultimately people caring about people.

Taking 9 participants, we used the Tobii X2-60 eye tracker and iMotion to capture eye movement on the screens with gaze mapping and the participant's human behavioral reaction when using the Verizon and AT&T websites. We conducted this competitive usability evaluation at the Texas A&M Behavioral Sciences Lab in College Station, Texas. We decided to use the Tobii X2-60 Eye Tracker because paired with iMotion it yields attuned empirical data that we can bridge with qualitative data. Emotional response is considered qualitative data, but by capturing empirical data of facial movements we can translate certain muscle movements to emotion. This software is capturing a live empathy map, gathering data on what the user does and how they feel.

Each user was tasked with locating the Iphone 11 Pro and adding it to cart without using the search bar. The reason we decided we wanted the user to locate the Iphone 11 Pro was because it isn't the newest model, so it isn't advertised on the first page, there are no quick links to it. The user has to look and filter devices to locate the Iphone 11 Pro, forcing them to explore the site and search. At the completion of the test the users were asked a series of questions including which site they preferred and if their experience on the site had an effect on which service they would sign up for. We looked aimley in the data for the time to first fixation on the Iphone 11 Pro utilizing gaze mapping to do so. The goal of this usability test was to gather data and exemplify that this method of usability testing can be used to introduce a new competitive edge to corporations. If a company can uncover the bottlenecks in their own applications or the

applications of their competitors and execute changes, their site or application will be more user centric, leading to happier customers.

SECTION V

USABILITY TEST

Looking at usability testing best practices along with the added emphasis of eye-tracking and behavioral response, we created the outline for our competitive moderated in person in lab usability evaluation as seen below. As defined earlier in the paper the study is competitive as we are determining how AT&T's site performs in relation to Verizon's

Site 1:
Verizon
https://www.verizon.com

Site 2:
AT&T
https://www.att.com

Prompt on screen:

Your goal is to locate the Iphone 11 Pro and add to cart without the usage of the search bar on either site.

Clarifying question:

What are you looking for?

a. Iphone 11
b. Iphone 11 Pro

c. Iphone 12

User proceeds to complete the task on both sites. Sites appear in a randomized order, different per user.

Debrief Questions: (google form)

1. Which service provider site was easier to use?

2. Why do you think it was easier to use?

3. Which site did you have to click more on to complete a prompted action?

4. Did you note a virtual assistant that appeared? How likely are you to use this?
5. Have you used one before, was your experience positive of negative?

6. Was it easier on one of the sites to compare plans and figure out which was the best for you?

7. Was it notably hard to complete an action on either of the 2 sites?

8. Was it challenging to find a button on either of the 2 sites?
9. Were you at any point unsure of yourself when completing an action?

10. Were you at any point unable to complete an action?

11. Did you get annoyed or anxious at any point trying to complete an action?

12. Which site did you find more satisfying to use?

13. Did one site feel more trustworthy than the other?

14. Do you think the website has any added influence to your choice when choosing a service provider?

15. Based on the website alone which service provider would you choose to sign on with?

SECTION VI

RESULTS

The results of the test showed that on average it took users 21.667 seconds to locate the Iphone 11 Pro on the AT&T site and an average of 111.111 seconds on the Verizon site, as seen in Fig. 5. The test also recorded the emotional response of the user. We were able to see when the user experienced anger, contempt, disgust, fear, joy, sadness, surprise, engagement, and valence. Engagement, surprise, and joy commonly occurred when the user had located the product or when they couldn't find the product as seen in Fig. 10 and 11. The iMotions emotion analyzer doesn't do well recording negative emotions, as more often than not the user's facial reaction do not reflect them, but we can see when a user is engaged or joyful moreso.

Time to first fixation:

	ATT	Verizon
Kr	19	113
Mad	29	92
Dom	17	114
Wi	21	164
Ta	18	43
Ha	29	130
Mi	17	68
Co	29	198
Ai	16	78
Avg	21.667	111.111

Fig. 5 User's time to first fixation on Iphone 11 Pro

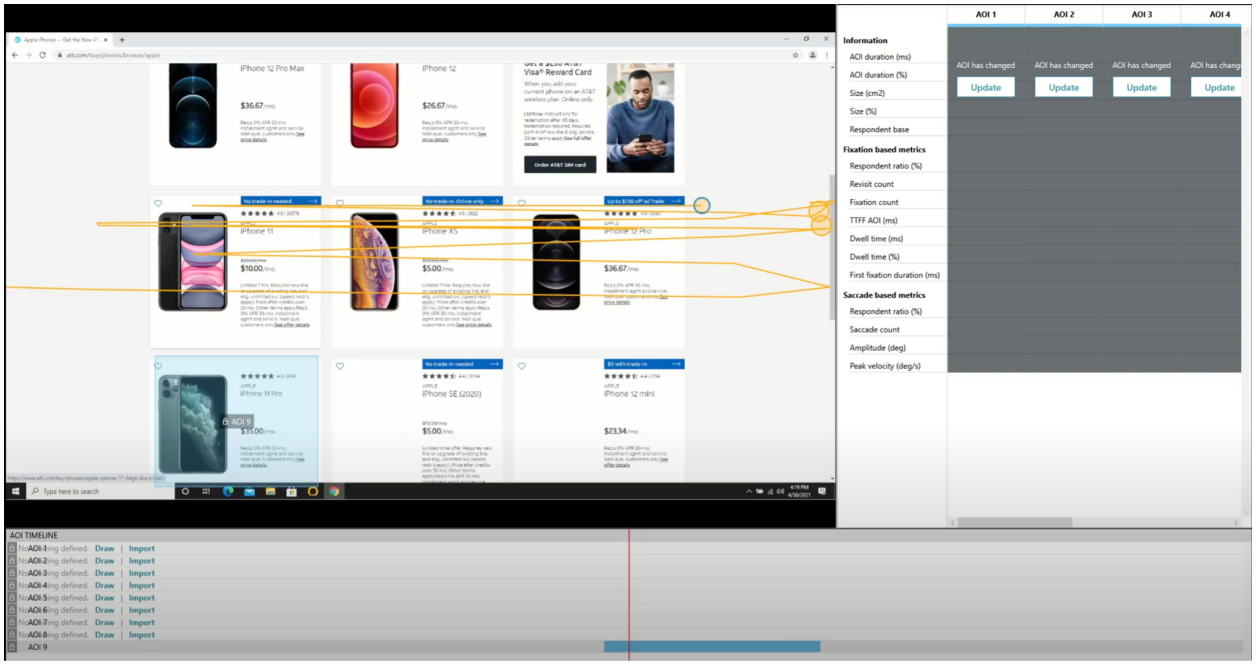

Fig. 6 Displays a gaze map, shows the tracking of the eyes as they read across the page looking for the Iphone 11 Pro on the AT&T site.

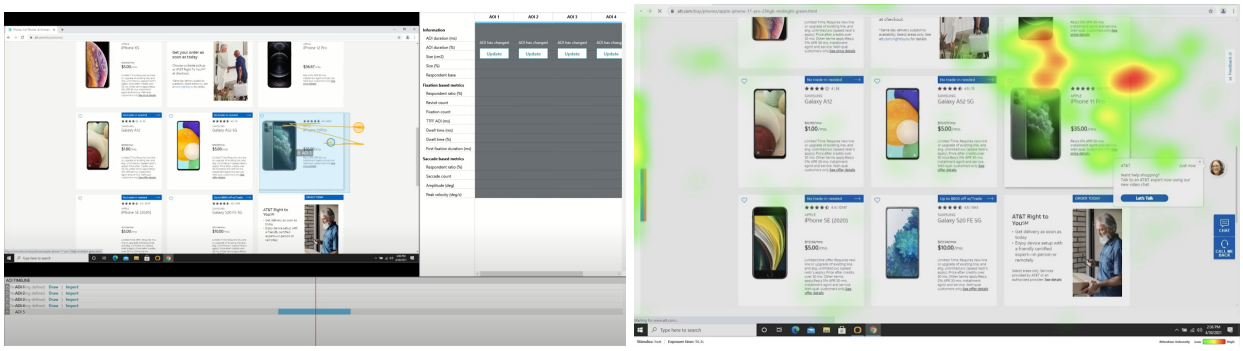

Fig. 7 Displays moment of first fixation on the Iphone 11 Pro on the AT&T site, you can see the eyes are focused on this one area. A heat map is also displayed her, validating the users point of fixation on the Iphone 11 Pro.

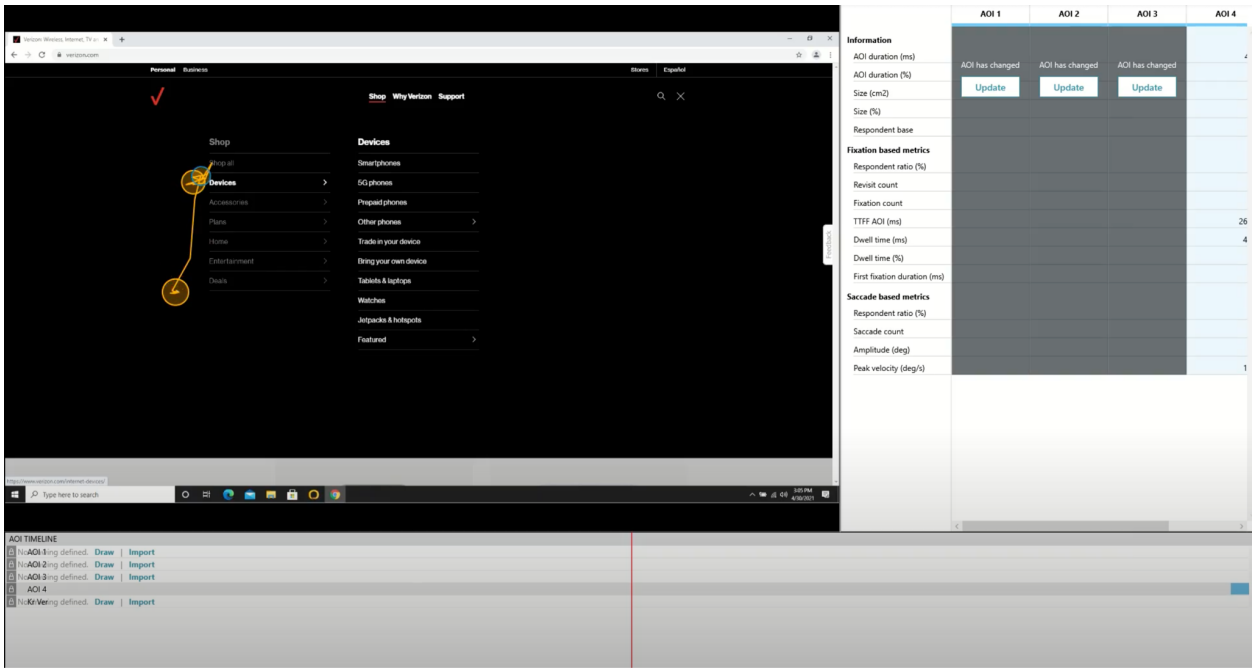

Fig. 8 Displays the first necessary interaction point on the Verizon site in which the user must decide what they are looking for and how to get there.

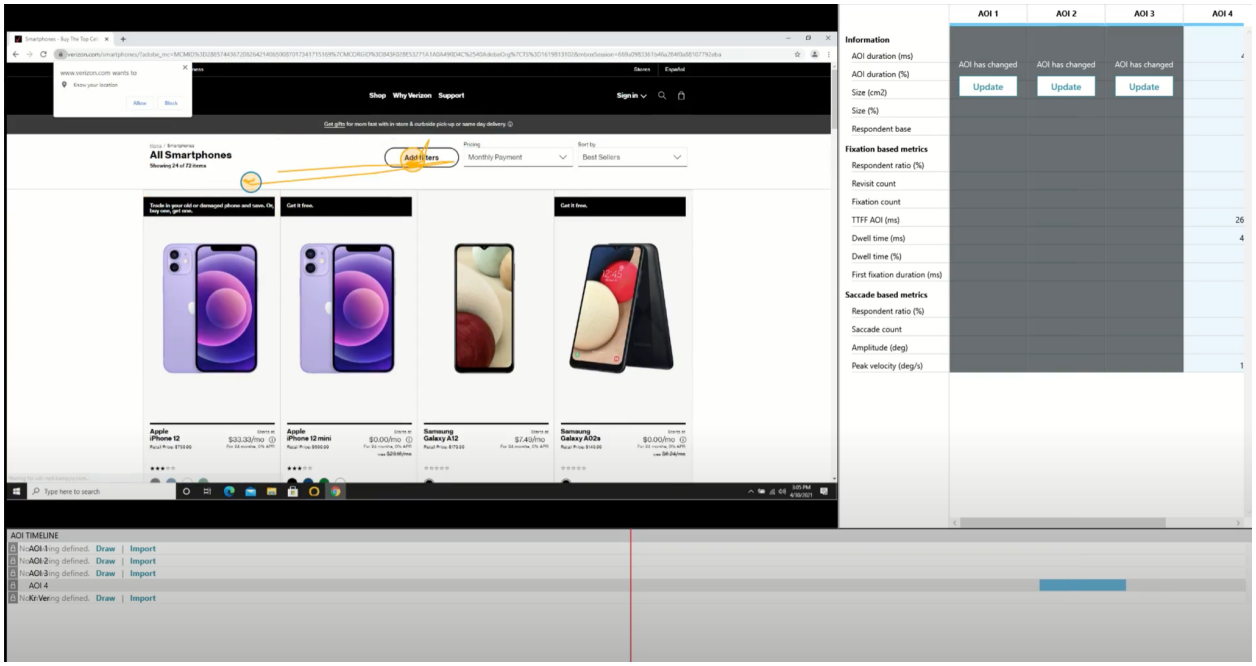

Fig. 9 Displays the user noticing the filter button, the second necessary interaction point for the user to complete in order to locate the Iphone 11 Pro on the Verizon site.

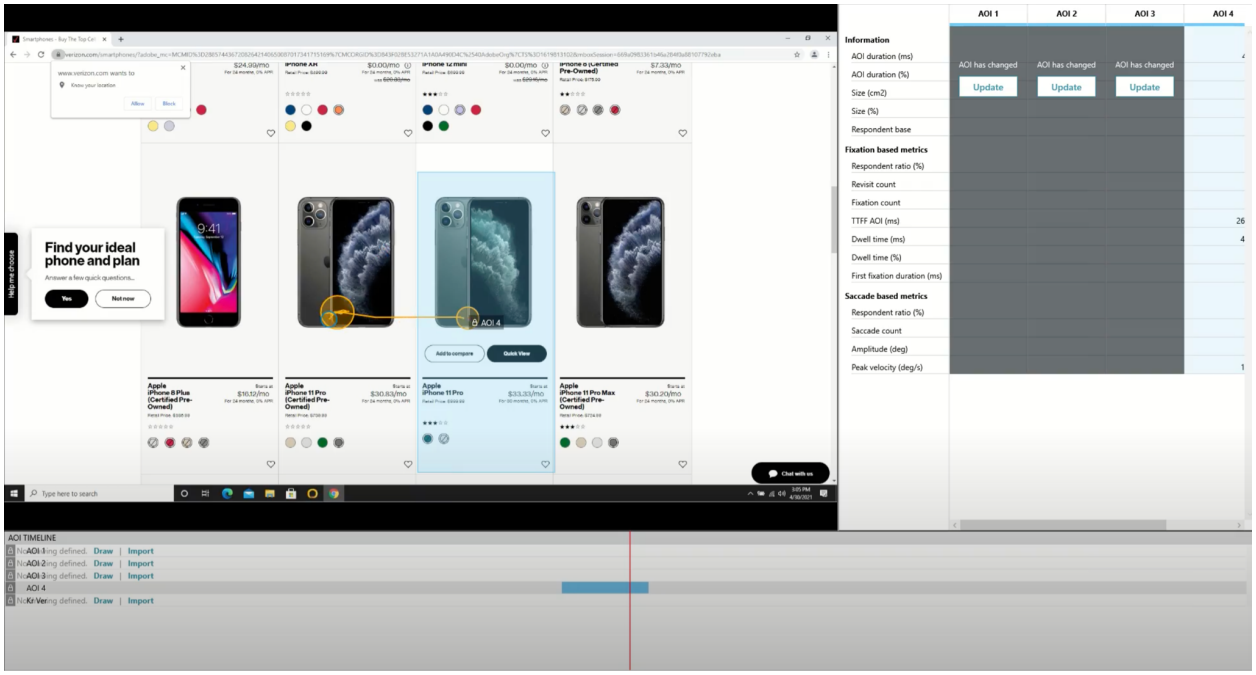

Fig. 10 Displays the user's moment of first fixation on the Iphone 11 Pro on the Verizon site.

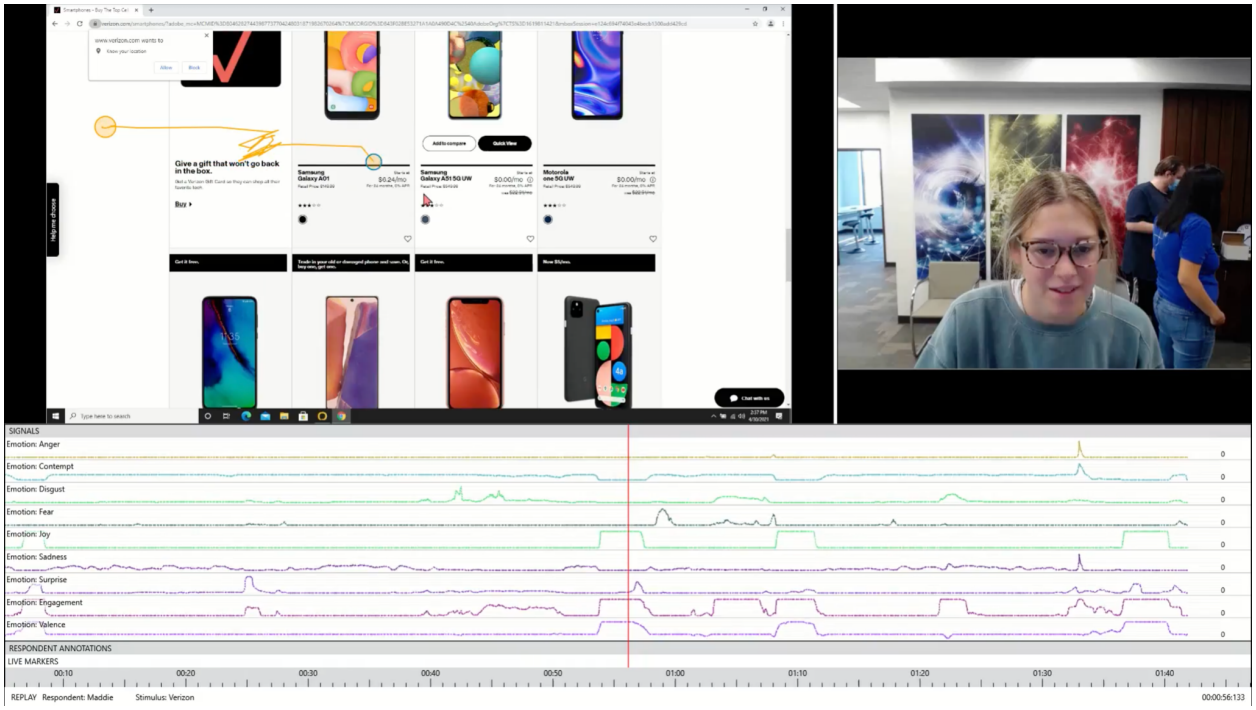

Fig. 11 Displays behavioral tracking of a user smiling, sending a spike in joy, engagement, and valence.

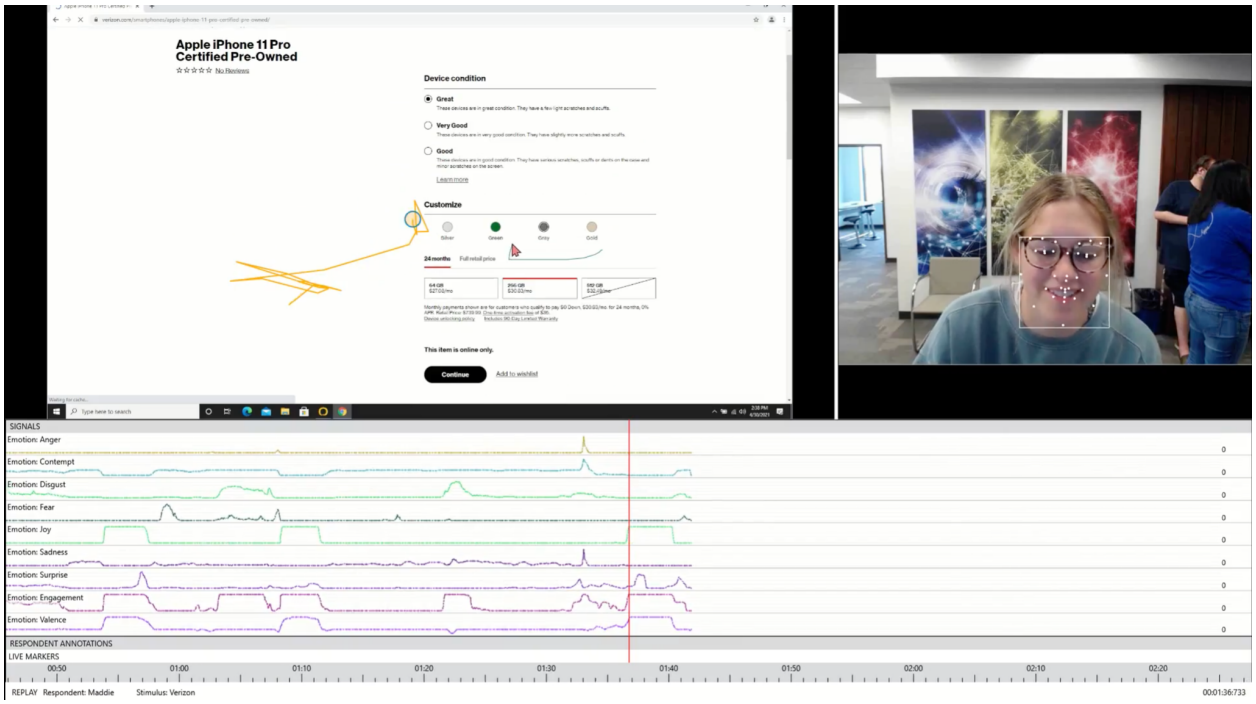

Fig. 12 Example of a moment when the user is smiling, displaying higher levels of engagement, joy, and valence.

User debriefs after the usability test were sent out in survey form. Six responses of the nine participants were received. Some of the more interesting results from the survey can be seen in Fig. 13, 15, 17, 18, 19, and 20.

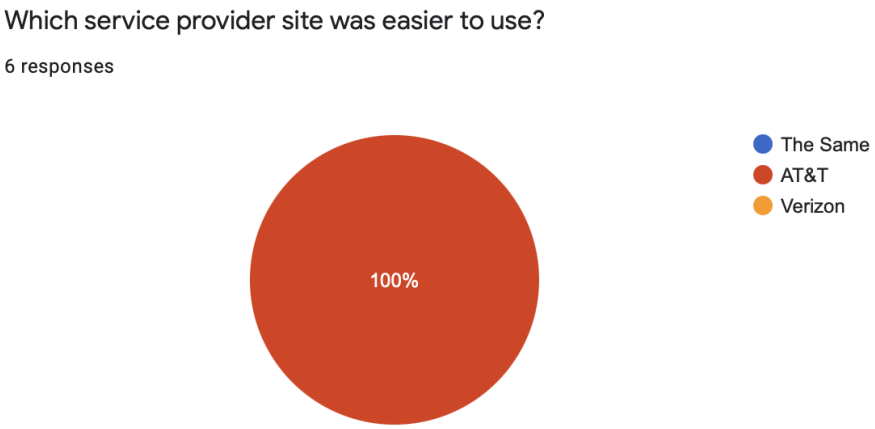

Fig. 13

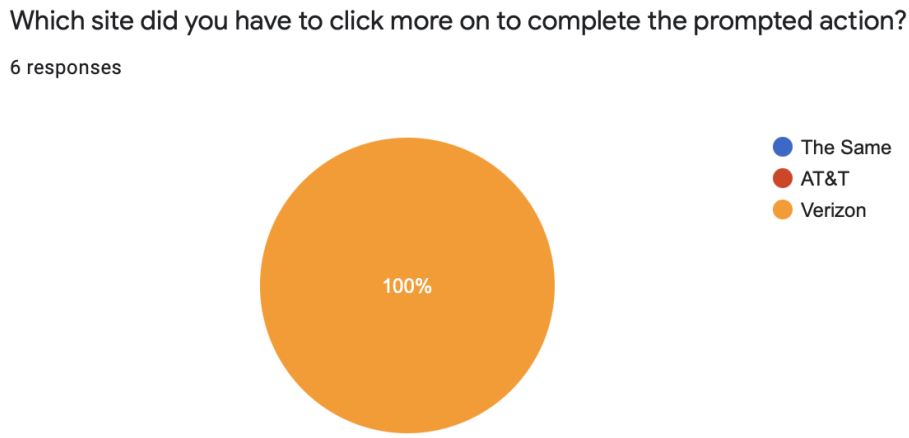

Fig. 14

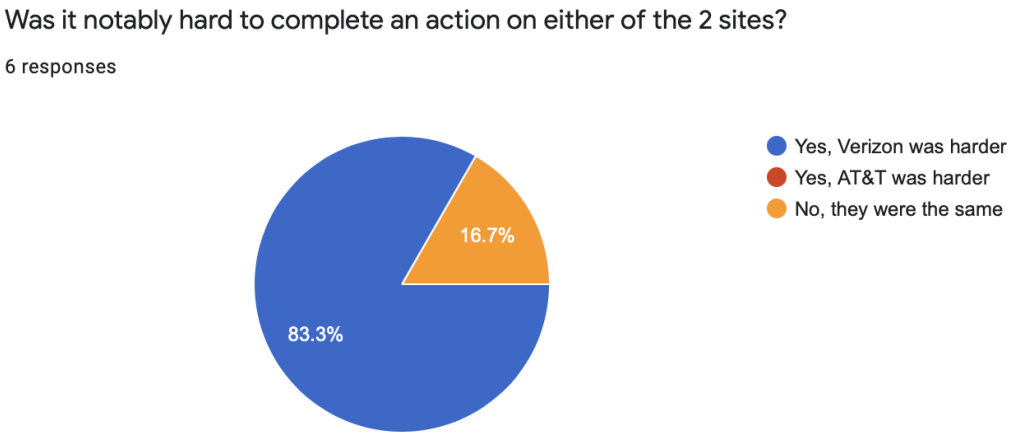

Fig. 15

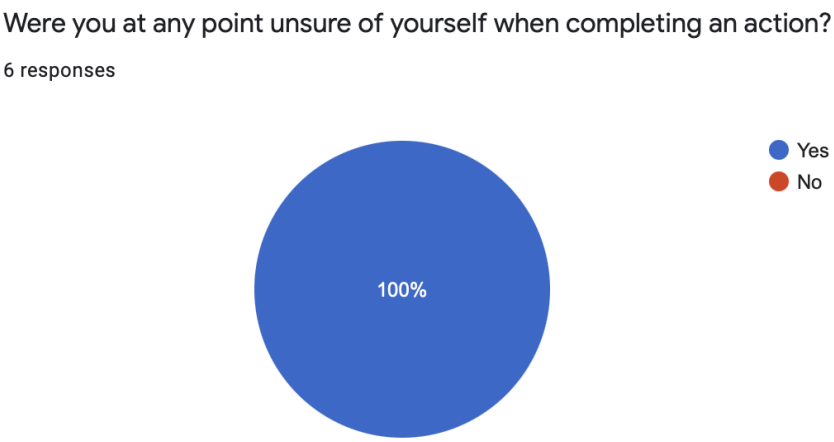

Fig. 16

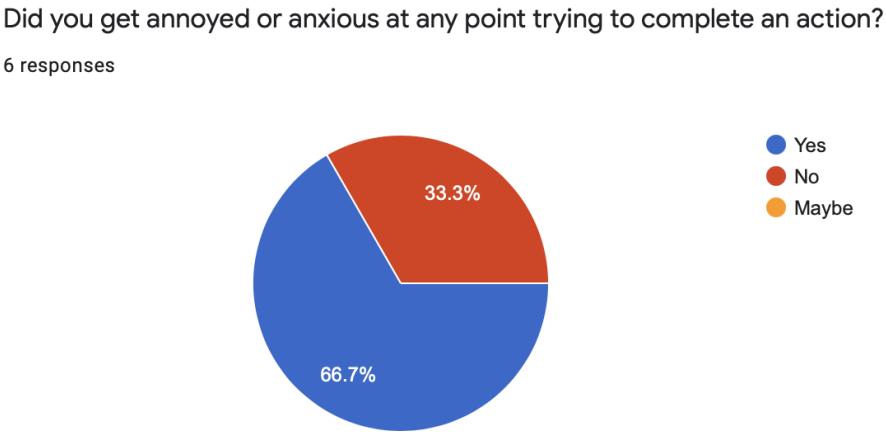
Did you get annoyed or anxious at any point trying to complete an action?
6 responses
33.3%
66.7%
Yes
No
Maybe

Fig. 17

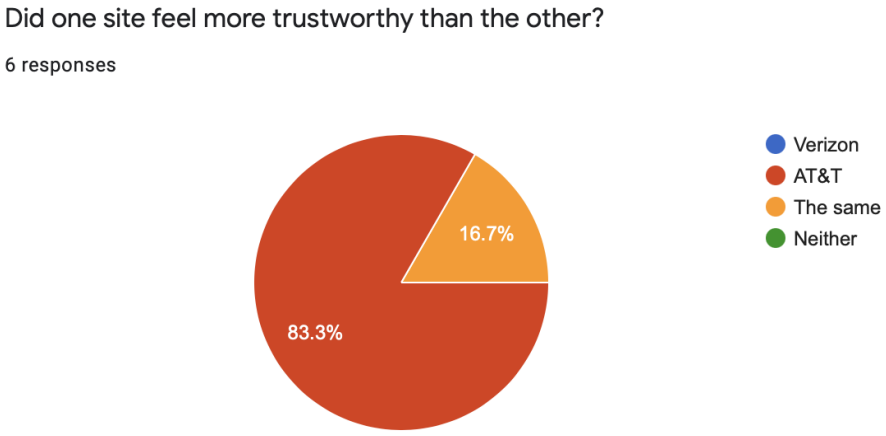
Did one site feel more trustworthy than the other?
6 responses
16.7%
83.3%
Verizon
AT&T
The same
Neither

Fig. 18

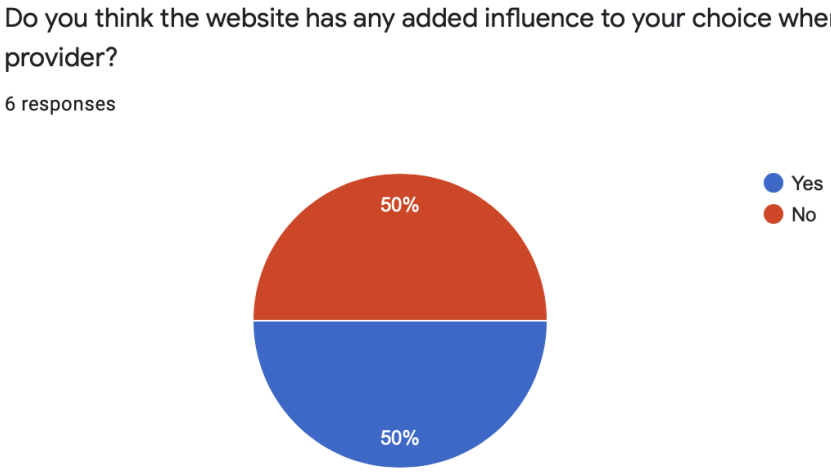
Do you think the website has any added influence to your choice when choosing a service provider?
6 responses
50%
50%
Yes
No

Fig. 19

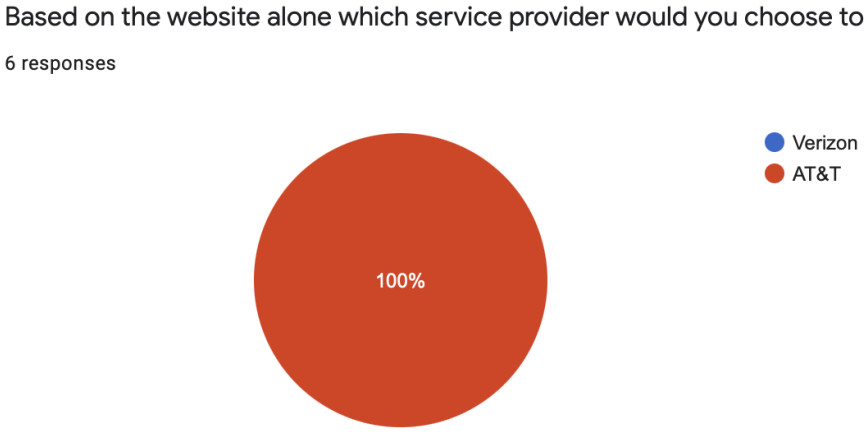
Based on the website alone which service provider would you choose to sign on with?
6 responses
100%
Verizon
AT&T

Fig. 20

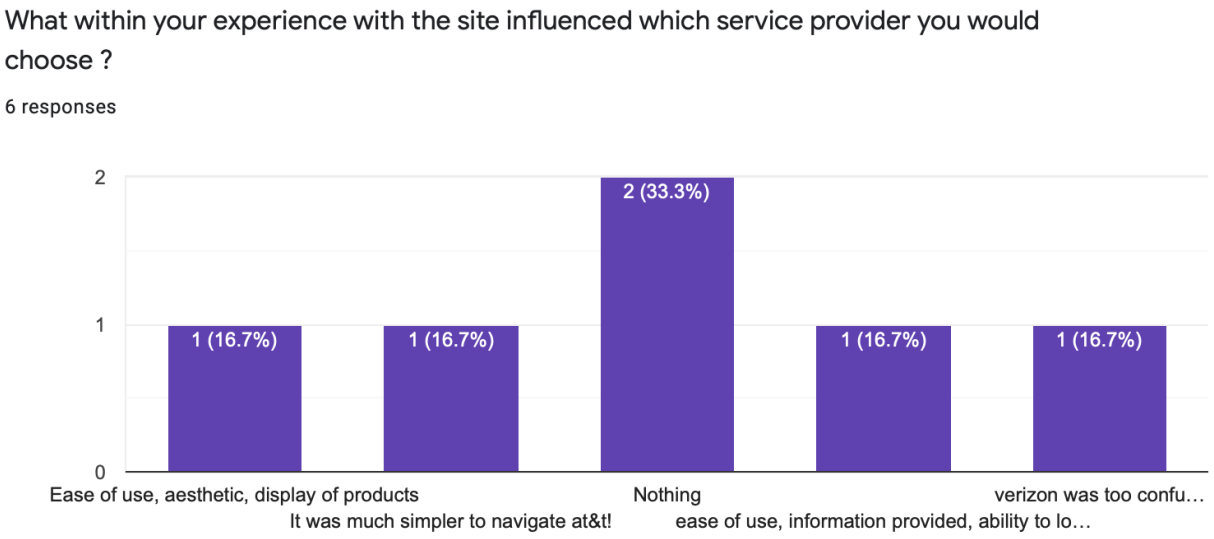

Fig. 21

SECTION VII

ANALYSIS

The data collected provides a basis for us to see the time to first fixation between the AT&T and Verizon sites had a very large disparity. The behavioral data collected showed points in the Verizon site where the user had to be more engaged to locate the Iphone 11 Pro. With the ability to watch the recorded usability test of each participant seeing where their areas of fixation went and what their behavioral reaction was, as they go through the website it is easy to pull out what the bottlenecks are. User's preferred the AT&T site to the Verizon site because it was easier to use, less clicks to complete an action, and more trustworthy. 50% of users said that the site had an added influence over which service provider they would sign up for and 100% said they would choose AT&T for reasoning ranging from "ease of use, aesthetic, display of products."

The perfect journey within the AT&T site consisted of the user clicking “Phones & Devices" when they initially come to the landing page and then simply scrolling down the page to locate the Iphone 11 Pro. The perfect journey within the Verizon site consisted of the user clicking “Shop,” then hovering over “Devices,” then clicking “Smartphones,” filtering the page to be only “Apple(21)” products, then scrolling to find the Iphone 11 Pro. With so many necessary clicks to complete the task of finding the Iphone 11 Pro on the Verizon site users had more chances to deviate, and many did. The “Add filters" button didn't catch the user’s eye and many would completely ignore it, whereas in AT&T when the users used the filter for Apple products it was easy to see and use. This was the main bottleneck in finding the Iphone 11 Pro. It wasn't with all the other smartphones, the smartphones had to be filtered to locate it, but the filter button wasn’t intuitive enough for the users.

Because of usability testing specifically within eye tracking and behavioral response I was able to deduct which site had the most ease of use and able to specifically identify problems within the usability of the site that inevitably could cause Verizon to lose customers due to their dissatisfaction with the website when located products or services.

SECTION VIII

REFLECTIONS

The evolution and recent emphasis placed on the importance of user experience has created a space in which usability testing is a necessary asset to remain user centric within design. The progression in eye tracking and behavioral tracking technology has provided a solution to gaining deeper insights into human behavior and interaction with a product. Seeing where the user’s eyes are drawn and looking into the behavioral reaction to complete a task

hones user centricity and makes it easier to create effective changes and solutions within design. Usability testing best practices with the implementation of eye tracking and behavioral tracking technology in a business environment introduces a way to gain a better understanding of the end user and a user's needs when interacting with products or services.

WORKS CITED

Bennett, Jessica. “Formative Vs Summative : The User Testing Battle.” *Usability Geek*, usabilitygeek.com/formative-vs-summative-the-user-testing-battle/.

Cooke, Lynne. “Is Eye Tracking the Next Step in Usability Testing?” *IEEE Xplore*, 2006, ieeexplore.ieee.org/abstract/document/4114166/authors#authors.

“Improving Your Research with Eye Tracking since 2001: Tobii Pro.” *Improving Your Research with Eye Tracking since 2001 | Tobii Pro*, 27 Apr. 2015, www.tobiipro.com/.

Jakob Nielsen Technical University of Denmark, et al. “Heuristic Evaluation of User Interfaces.” *Heuristic Evaluation of User Interfaces | Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1 Mar. 1990, dl.acm.org/doi/abs/10.1145/97243.97281.

Manhartsberger, Martina, and Norbert Zellhofer. “Eye Tracking in Usability Research: What Users Really See.” *Empowering Software Quality: How Can Usability Engineering Reach These Goals?*, 2005, www.usability.at/download/EyetrackinginUsability.pdf.

Pro, Tobii. “Spectrum: Most Advanced Eye Tracking Platform: Find out More.” *Spectrum: Most Advanced Eye Tracking Platform | Find out More*, 10 Nov. 2016, www.tobiipro.com/product-listing/tobii-pro-spectrum/.

Sears, Andrew, and Julie A. Jacko. *Human-Computer Interaction: Development Process*. CRC Press/Taylor & Francis Group, 2009.

Thyvalikakath, Thankam Paul, et al. "Comparative Study of Heuristic Evaluation and Usability Testing Methods." *Studies in Health Technology and Informatics*, U.S. National Library of Medicine, 2009, www.ncbi.nlm.nih.gov/pmc/articles/PMC2736678/.

Tobii. "Tobii Pro Glasses 2 User's Manual: Head Unit Recording Unit Controller Prescription Lenses." *TobiiPro*, Tobii AB, Nov. 2020, www.tobiipro.com/siteassets/tobii-pro/user-manuals/tobii-pro-glasses-2-user-manual.pdf.

Tobii. "User's Manual Tobii X2-60 Eye Tracker." *TobiiPro*, Tobii Technology AB, June 2014, www.tobiipro.com/siteassets/tobii-pro/user-manuals/tobii-pro-x2-60-eye-tracker-user-manual.pdf/?v=1.0.3.

World Leaders in Research-Based User Experience. "Benchmarking UX: Tracking Metrics." *Nielsen Norman Group*, www.nngroup.com/articles/benchmarking-ux/.

World Leaders in Research-Based User Experience. "Competitive Usability Evaluations: Definition." *Nielsen Norman Group*, www.nngroup.com/articles/competitive-usability-evaluations/.

World Leaders in Research-Based User Experience. “How to Measure Learnability of a User Interface.” *Nielsen Norman Group*, www.nngroup.com/articles/measure-learnability/.

Written by David Leggett Chief Design Off, et al. “A Brief History of Eye-Tracking: UX Booth.” *UX Booth A Brief History of EyeTracking Comments*, www.uxbooth.com/articles/a-brief-history-of-eye-tracking/.