MSc in Computing - Team Project

User Evaluation Report

Dark Pattern Detector

Lan Zeng - D19124089

Alan Dowley - D19127378

Quanwei Sun - D19126879

Jordan Donnelly - D20125012

Yufei Su - D18130277

Yunpeng Liu - D20123758

Andrea Curly
Damian Gordan
Paul Kelly

18/11/2021

Table of Contents

1. Introduction

User evaluation is an integral part to any product or system design. It allows designers to get a better understanding of current design issues and helps designers to reflect on the needs and wants of the users. All while, attempting to shine light on the context, stimuli, environment factors, time factors and a lot more (IshΔn, 2019). User evaluation and usability testing can be done in many different forms, from a simple questionnaire about the product to actually letting the users use it and gathering their reactions and thoughts throughout the process.

In this report we first describe what user evaluations we have done on our system and the different types of user evaluations used in each time. Next, talk about the design of the user evaluation experiments, going through how we will set them up, how data will be collected, the users and how data will be analysed. The results of each user's evaluations will follow on from that, where we show our quantitative and qualitative results and discuss some of our findings as well. Lastly, will be our conclusion, which is an overview of our findings as well as thoughts on what we would do differently if we were to do more user evaluation.

2. Proposed Hypothesis

Generally, we conducted two user evaluations. The first and second evaluation processes are similar, using both Cognitive Walkthrough and questionnaires. However, in the first evaluation, we did not consider the expert users. In the second experiment, we improved the evaluation methodology to include expert reviews and AB tests.

The questions below are those used in the evaluation, and we will present them from two aspects: Cognitive Walkthrough and Questionnaire.

2.1 Cognitive Walkthrough Questions

For user evaluation, three different testing methodologies are applied. Cognitive walkthrough, questionnaires, and heuristic evaluation. In terms of Cognitive walkthrough, we first put forward three questions to the observers, which are as follows:

1. Is the user able to locate the area to complete the task?
2. Is the user able to make the correct action without assistance?
3. Does the user associate the correct action with the correct expected result?

Observers collected data on these three questions as a basis for evaluating whether the user could complete the task.

In the questionnaire evaluation, the questions we need to collect are divided into two categories, one is open-ended qualitative questions and the other is quantitative questions, using a Likert Scale or star rating. Using qualitative questions helps to assess the whole system more comprehensively, while quantitative data can show the severity of the problem.

The test subjects were five ordinary users and two experts, who first tested the application via 8 user tasks and then were presented with 12 open-ended qualitative quantitative questions. The user tasks are listed below:

User Task:

1. How would you find more information on Dark Patterns?
2. How would you detect Dark patterns on a webpage?
3. How would you report a dark pattern on a webpage?
4. How do you turn on/off Autoscan?
5. How do you turn on/off the detection of certain types of patterns?
6. How would you locate highlighted patterns on webpages?
7. Can you find what each icon represents?
8. Where do you find information about checkboxes and page location, in the extension?

2.2 Questionnaire Qualitative question

Qualitative questions directly identify major usability issues in the interface.

1. What is your opinion on the design and colours used within the extension?
2. Would you change the layout or design of any part of the extension if you had the chance?
3. Do you understand what each icon means and how it relates to the pattern type?
4. Do you understand what needs to be filled out in each section of the report page? why or why not?
5. Do you understand what each feature in the extension does?
6. Is there too much or too little information displayed in the extension?
7. What do you think is the best thing about the extension?
8. What do you think is the worst thing about the extension?
9. Do you have any suggestion how the extension could be made more user friendly?
10. What other features would you like to see in the extension?
11. Do you find this extension useful?
12. With the help of this extension, do you know what a Dark Pattern is now?

2.3 Questionnaire Quantitative questions

1. The text in the extension is easy to read and understand. (Likert Scale 1-5)
2. The colour used across the extensions UI, make it easy to read information in the extension. (Likert Scale 1-5)
3. The extension has the same feel throughout its entirety. (Likert Scale 1-5)
4. Detection data presented in the extension is always in the same manner across different webpages. (Likert Scale 1-5)
5. The amount of information on a single page of extension is appropriate (i.e., “About” page, “Result Page”, “Report” Page, and “Settings” page). (Likert Scale 1-5)
6. Rate the quality of information displayed in the UI of the extension. (Star Rating 1-5).
7. Rate the relevance of the information displayed by the extension. (Star Rating 1-5).
8. Rate consistency of extension design style. (Star Rating 1-5)
9. Rate the visual clutter of the user interface of extension. (Star Rating 1-5).

According to Dewi et al., 2020. Cognitive drills are used to capture data about the implementation of extended functionality, time consumption, and so on. Questionnaires were used to obtain user satisfaction, ease of use, readability, accessibility, visual perception, data consistency, etc. User experience data from cognitive exercises and user evaluation questionnaires can be combined to suggest improvements. Heuristic Evaluation techniques discover usability problems in design from an expert's perspective.

3. Experimental Method

3.1 Overview

The user experience evaluation of the Dark Pattern Extension is an evaluation study that integrates qualitative and quantitative data and adopts a mixed method. The mixed method is a combination of quantitative and qualitative methods, in which the collection of experimental data is divided into the following stages:

1. Determination of user task and questionnaire.
2. Establishment of experimental steps
3. Data collection
4. Data analysis

5. Put forward suggestions for improvement. At this stage, conclusions are drawn based on data analysis results and suggestions are put forward for improvement of extension application interface.
6. Implement the change plan.
7. Feedback results are obtained from AB testing.

User Task Evaluation：

Test the functionality of the extension application using a cognitive walkthrough approach. In this approach, users are provided with information about performing pre-set user tasks, and the tester records user responses and fills out forms. These tasks are the actual actions of the user when using the plug-in.

At the end of each task, the recorder answers 3 default questions and completes the recording form. The questions are:

1. Is the user able to locate the area to complete the task?
2. Is the user able to make the correct action without assistance?
3. Does the user associate the correct action with the correct expected result?

If the answer to all four questions is yes, the mission is called a "success story." If the answer to one of the four questions is no, the mission is called a "failure story" of that mission. Finally, performance was evaluated according to the task completion rate table provided by the observer.

User Experience Questionnaire Evaluation：

There are 9 questions to measure user experience from six aspects: Readability, Data correlation, Accessibility, Visual aesthetics, Visual overload, and satisfaction. The questionnaires will be distributed to the participants after the Cognitive Walkthrough. The questionnaire consists of the above 8 questions. The score for each availability scale consists of the interval scale score. Participants' responses to the questionnaire were submitted and stored in a Google form.

3.2 Data Collection

During the survey, we would ask users or record their behaviour or reaction from three directions as mentioned in Cognitive Walkthrough questions above. Cognitive Walkthrough are used to evaluate usability of a product (Dalrymple, 2018), by focusing on the level of easiness and the length of time taken for users to complete their tasks.

	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7	Task 8
	Task Completion / Time Taken							
User 1	y/y/y/5s	y/y/y/6s	y/y/y/15s	y/y/y/9s	y/y/y/5s	s/s/s/15s	s/s/s/5s	y/y/y/5s
User 2	y/yn/3s	y/y/n/3s	y/y/y/13s	n/n/n/23s	n/n/n/30s	s/s/s/3s	y/n/n/13s	y/n/n/3s
User 3	y/y/y/7s	y/y/y/8s	y/y/y/7s	y/y/y/7s	F/17s	s/s/s/7s	n/n/n/7s	y/y/y/17s

Table 1 - Data collection for Cognitive Walkthrough (3 Answers & Time taken)

In this evaluation test, there will be mixed quantitative and qualitative questions in our survey. The first nine questions are quantitative questions. And the last are qualitative questions. The feedback, suggestions and issues given by users from the questionnaire will all be qualitative. We use Likert scale of 5-point agreement scale for users to score: (1) Strongly disagree; (2) Disagree; (3) Neither agree nor disagree; (4) Agree; (5) Strongly agree. The metrics which are measured in questionnaire are shown below:

Readability score: It tells how easy for users to read the information in our Dark Pattern detection extension.

Data relevance: According to Hes (2017), data relevance is the extent to which data answers or gives insight into the question of the target user. It measures if we provide relevant dark pattern data which the user requires.

Accessibility: Accessibility allows users of all abilities to understand, use and interact with our extension (Kuar,2018). It should be made for all users with different background to use our product.

Consistency: Consistency needs to achieve both in terms of visual experience and design. It's important to provide users consistent or close reexperience in our extension.

Visual Weight:

Visual overload: Options and colourful icons, chart may overload user's visual sensor and confuse them.

Visual aesthetics: A successful user interface will make users enjoy using our extension. It defines a design's pleasing qualities and make them more tolerant of usability issues. (International Design Foundation)

The relationship between evaluation indicators and question index is shown as the table below:

Evaluation Indicators	Question Index
Readability score	Q1
Data relevance	Q6 and Q7
Accessibility	Q2
Consistency	Q3, Q4, and Q8
Visual weight	Q5 and Q9

Table 2 - Relationship between evaluation indicators and question index

3.3 Selected Subjects

It is desired that the system can be used by all sorts of people ranging from Dark Pattern experts, everyday shoppers, those who have fears of being deceived etc. Due to this, we plan to evaluate 3 kinds of users for our system

- People similar to personas: These are users that share traits with our created personas. Selected users, regular online shoppers.
- Common person: This is no particular person, but someone who does not share many traits of our created personas. Relatives, friends, and classmates.
- Experts: These are people who are considered experts in the field, whether it be experts in UX or Dark Patterns. Lecturers.

Without evaluating such users, it would be immensely difficult or near impossible to reach a sufficient level of usability throughout our system that is desired.

These are the representative samples because they match our target user groups. The idea is to allow as many people as possible to use our system and for it to benefit them, regardless of if they have knowledge of dark patterns or not. Cognitive walkthrough

was used for evaluation in the first half of development. The main issues that the users listed were as follows:

- There is no information on what a dark pattern is.
- The highlighting feature was too dull to see.
- The users had to search to find the highlighted dark pattern.

These tests were conducted with people who are similar to our personas and common people. The evaluation of these results allowed us to change the UI of the chrome extension to become more user friendly. We added a page dedicated to informing the user as to what a dark pattern is. We changed the colour and border of the highlighted dark patterns to be clearer and added a feature that would allow the user to be directed to any of the selected dark patterns.

In the second half of the development, we would conduct surveys and questionnaires as well as cognitive walkthroughs with the 3 user groups listed above. The testers are sourced through relatives, friends, classmates, lecturers, housemates, and those who are frequent online shoppers.

3.4 Data Analysis

3.4.1 Data Analysis of Cognitive Walkthrough

The data collected from the Cognitive Walkthrough tasks are in 2 parts, and the measuring criteria is shown in Table 3.

1. The task completion status, according to the response of the 3 Cognitive Walkthrough questions mentioned in Section 2.1.
2. Time taken for each task.

Metric Measured	Measuring Criteria (Success)	Measuring Criteria (Failure)
Task Completion Tasks are either assigned “Success” or “Failure” for each user based on whether the user is able to complete the task.	All of the responses towards the 3 Cognitive Walkthrough questions are “Yes”.	Any of the responses towards the 3 Cognitive Walkthrough questions is “No”.
Time Taken	≤10s	>10s

Table 3 - Measuring Criteria of Cognitive Walkthrough

Table 4 shows an example of data collected from Cognitive Walkthrough, manipulated according to the measuring metrics. The example table shows the result of 2 users doing 8 tasks in Cognitive Walkthrough, to test the extension functionalities. Each task

is assigned two values, representing 'Task Completion Status' and 'Time Taken'. For example, "S/5s" means the task is completed successfully and the time taken for the task to complete is 5 seconds.

	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7	Task 8
	Task Completion / Time Taken							
User 1	S/S	S/S	S/F	S/S	S/S	S/F	S/S	S/S
User 2	F/S	F/S	S/F	F/F	F/F	S/S	F/F	F/S
User 3	S/S	S/S	S/S	S/S	F/F	S/S	F/S	S/F
Task Score	66.7%	66.7%	100%	66.7%	33.3%	100%	33.3%	66.7%
Time Score	100%	100%	33.3%	66.7%	33.3%	66.7%	66.7%	33.3%

Table 4 - Example of data collected from Cognitive Walkthrough

'S' = 'Success', 'F' = 'Failure'

As shown in the example table, 48 attempts (24 Task Completion Status attempts + 24 Time Taken attempts) in total are observed to perform the tasks. Among all the attempts, 32 were successful, and 16 were failed. All the success points contribute to the success rate and all the failure points contribute to error rate.

According to this example, the success rate of Cognitive Walkthrough would be: 32/48 = 66.7%. The error rate of Cognitive Walkthrough would be: 1- 66.7% = 33.3%.

3.4.2 Data Analysis of Questionnaire

The data collected from the responses of quantitative questions in the questionnaires, which was presented to users after the Cognitive Walkthrough, is analysed according to 5 metrics, Readability, data relevance, accessibility, consistency, and Visual Weight. Each of the 9 quantitative questions is mapped with a certain metric as shown in Table 5. For each quantitative question in the questionnaire, the Likert scale with 5-point agreement scale is applied.

Metric Measured	Question Index	Measuring Criteria (Success)	Measuring Criteria (Failure)
		Calculated based on the average score of the related questions	

Readability	Q1	≥4	<4
Data Relevance	Q6 and Q7	≥4	<4
Accessibility	Q2	≥4	<4
Consistency	Q3, Q4 and Q8	≥4	<4
Visual Weight	Q5 and Q9	≥4	<4

Table 5 - Measuring Criteria of Questionnaire (Quantitative Questions)

Table 6 shows an example of data collected from the questionnaire, each metric parameter is assigned 2 values, one is the Success Status and the other is the average score of the parameter. For example, "F (2)" for 'Data Relevance' for 'User 2' means the average score of question 6 and question 7 is 2, and this is defined as a 'Failure' according to the measuring criteria in Table 4.

	Readability (Q1)	Data Relevance (Q6+Q7)	Accessibility (Q2)	Consistency (Q3+Q4+Q8)	Visual Weight (Q5+Q9)
User 1	F(2)	F(1)	F(3)	F(2)	F(1)
User 2	F(1)	F(2)	S(4)	F(3)	F(3)
User 3	F(4)	F(1)	F(3)	F(1)	S(5)
Average Score	2.3	1.3	3.3	2	3
Success Score	33.3%	0%	100%	33.3%	66.7%

Table 6 - Example of data collected from Questionnaire (Quantitative Questions)

'S' = 'Success', 'F' = 'Failure'

As shown in the example table, 15 attempts in total are organised according to the measuring criteria. Among all the attempts, 2 were successful, and 13 were failed. All the success points contribute to the success rate and all the failure points contribute to error rate.

According to this example, the success rate of Questionnaire would be: 2/15 = 13.3%. The error rate of Cognitive Walkthrough would be: 1 - 13.3% = 86.7%.

The **final success rate** is the average of the success rate of Cognitive Walkthrough and the success rate of Questionnaire, which would be (66.7% + 13.3%) / 2 = 40%, and the final error rate is 1 - success rate = 60%.

3.4.3 Data Analysis of Think Aloud

Data included in the Think Aloud are:

1. Qualitative results from the questionnaire
2. Issues the users meet during the Cognitive Walkthrough
3. Feedback and suggestion users gave during and after the Cognitive Walkthrough

3.5 Practical Setup

1. The first step of the experiment is to find users that represent our user personas created early in the project. Having a good selection of different types of users, would give us the most helpful feedback.
2. The users doing the experiment could either use a group member's computer or install it onto the personal computer, with the help of the group member carrying out the experiment.

3. Each experiment will then be carried out online, on one or more types of shopping websites, such as Amazon.com or Outfithustlers.com. The experiment can be carried out at any location, provided there is internet access.

4. Each user will be given a list of tasks that they are too complete using the extension. No other help will be given unless they request it. If help is requested, the task they are on and the help they asked for will be recorded by the tester. The user will work through each of the tasks provided and the tester will record how long it took them to complete each one, their reaction towards it, any obstacles they faced and any comments they made along the way.

5. Once the user completes each of the tasks, they will then be asked to fill out an online questionnaire about the experience they just had using the extension. This questionnaire is designed to give us both quantitative and qualitative data, from the user's feedback.

4. Results of User Evaluation

4.1 Results of User Evaluation with Extension Version 1

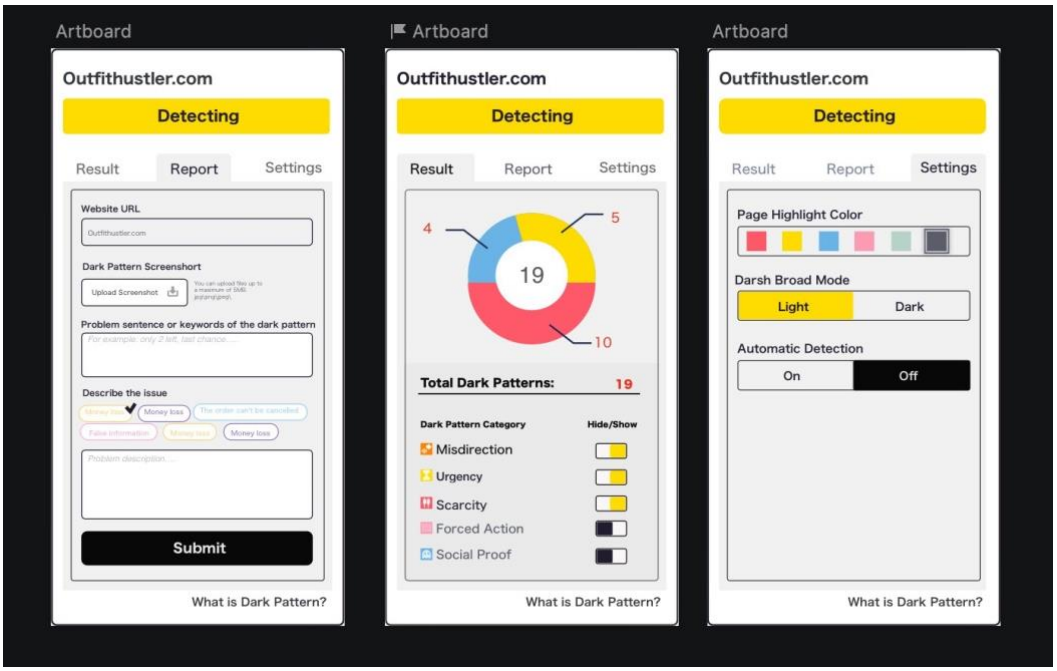


Figure 1 - Layout of Extension Version 1

4.1.1 Cognitive Walkthrough

For the Cognitive Walkthrough for extension version 1, 6 tasks were conducted by users, mostly the same as the 8 tasks in Version 2, only removing 2 tasks that was developed with Version 1. The tasks are shown as below:

1. How would you find more information on Dark Patterns?
2. How would you detect Dark patterns on a webpage?
3. How would you report a dark pattern on a webpage?
4. How do you turn on/off the detection of certain types of patterns?
5. How would you locate highlighted patterns on webpages?
6. Can you find what each icon represents?

	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6
	Task Completion / Time Taken					
User 1	F/F	S/S	F/F	S/S	F/F	S/S
User 2	S/S	S/S	S/F	F/F	F/F	S/S

User 3	F/F	S/S	F/F	S/S	F/F	S/S
User 4	S/S	S/S	F/F	S/S	F/F	F/F
User 5	F/S	S/S	S/S	F/F	F/F	S/F
User 6	F/S	S/S	F/F	S/S	F/F	F/F
User 7	F/S	S/S	F/F	S/S	F/F	S/S
User 8	S/S	S/S	S/S	S/S	F/F	S/F
Task Score	37.5%	100%	37.5%	75%	0%	75%
Time Score	75%	100%	25%	75%	0%	50%

Table 7 - Performance results obtained through Cognitive Walkthrough

From the result of the tasks, we could see at least 3 functionalities need to be improved, the issues are as below:

1. "Introduction" of Dark Patterns on the extension is not enough, users found it hard to see the link to the website. (Task 1)
2. The "Report" function on the extension, firstly the name itself is misleading to the users, should change it into something like "Report a Dark Pattern", secondly users found it difficult to understand what should be filled in in each field. (Task 3)
3. Most of the users can't locate the highlighted patterns on the webpage at all. (Task 5)

Performance results: 96 responses in total (48 Task Completion Status responses + 48 Time Taken responses) are organised according to the measuring criteria. Among all the responses, 52 were successful, and 44 were failed. All the success points contribute to the success rate and all the failure points contribute to error rate.

The success rate of Cognitive Walkthrough for Version 2 would be: 52/96 = 54.2%. The error rate of Cognitive Walkthrough would be: 1- 54.2% = 45.8%

4.1.2 Questionnaire (Quantitative)

For the questionnaire, 9 quantitative questions are presented to users, and 5 metric parameters are used for result evaluation, the results are shown in Table 8.

	Readability (Q1)	Data Relevance (Q6+Q7)	Accessibility (Q2)	Consistency (Q3+Q4+Q8)	Visual Weight (Q5+Q9)

User 1	S(4)	F(3)	S(4)	F(3.3)	F(3.5)
User 2	F(3)	S(4)	S(4)	S(4.3)	S(4)
User 3	S(4)	F(2.5)	F(3)	S(4)	S(4)
User 4	F(2)	F(2.5)	F(3)	S(4.6)	F(3)
User 5	F(2)	F(2)	S(4)	S(4)	F(3.5)
User 6	S(4)	F(2.5)	S(4)	F(3.3)	S(4)
User 7	F(3)	F(2)	S(5)	F(3.3)	F(3.5)
User 8	F(3)	F(1.5)	S(5)	F(3.3)	F(3)
Average Score	3.1	2.5	4	3.8	3.6
Success Score	37.5%	12.5%	75%	50%	37.5%

Table 8 - Results obtained from questionnaire (quantitative)

From the result of the questionnaire, most of the designs need to be adjusted to improve the usability, especially data relevance:

1. Data Relevance needs to be improved the most, what we presented on the extension UI doesn't really match what users expected.
2. The link to the website for more information and the text content in the extension need to be adjusted to improve readability.
3. The overall visual design is not eye-catching enough to attract users.

Questionnaire results: 40 responses in total are organised according to the measuring criteria. Among all the responses, 17 were successful, and 23 were failed. All the success points contribute to the success rate and all the failure points contribute to error rate.

The success rate of Cognitive Walkthrough for Version 2 would be: 17/40 = 42.5%. The error rate of Cognitive Walkthrough would be: 1- 42.5% = 57.5%.

The **final success rate** is the average of the success rate of Cognitive Walkthrough and the success rate of Questionnaire, which would be (54.2% + 42.5%) / 2 = 48.4%, and the final error rate is 1 - success rate = 51.6%.

4.1.3 Think Aloud

Data (in Table 9) included in the Think Aloud are:

1. Qualitative results from the questionnaire
2. Issues the users meet during the Cognitive Walkthrough

3. Feedback and suggestion users gave during and after the Cognitive Walkthrough

	Issues
Functionality	Needs some prompt beside the pie chart for explanation, not hover effect
	Hard to connect pie chart with the detected result
	What is this extension used for? Add the introduction about dark pattern
	Add a function that makes the detection to be automatic
Layout	Highlighting icon colour is too light to observe on the webpage.
	Hard to connect Highlight icons with the information's on the side window together
	Put a number to the pie chart for explanation
	Pie chart is not important, don care about it
Content	Question 3 and Question 4 on the report page is not clear
	"What is the Dark Pattern" looks not clear and not clickable
	The category title is too small make it bigger
	What do those small icons mean? User doesn't understand
	Confusing about the Icon colour on pie chart
	Not sure what does the highlighting area do
	"Number of total Dark Pattern" on the result page needs to be bigger

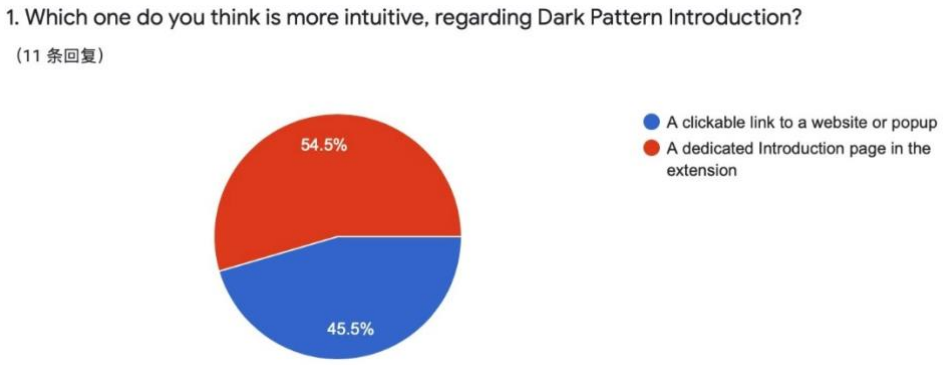
Table 9 - Results obtained from Think Aloud

4.2 Results of AB Test

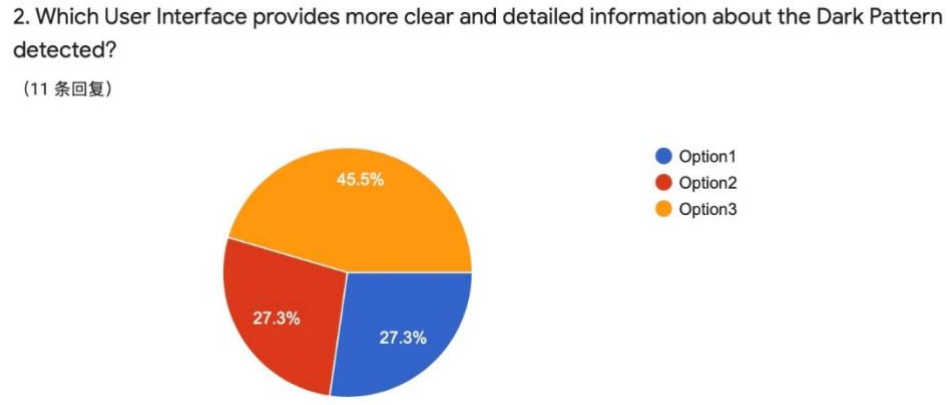
The AB test (Survey content is attached in Appendix) is conducted by 11 participants and the results are organised from the total of 11 participants. Among the responses, the first 6 responses (Group 1) are collected from the users who haven't used our

extension at all, and the last 5 responses (Group 2) are collected from the users who have used our extension and went through the Cognitive Walkthrough with us.

1. Among all the responses, 16.7% of the Group 1 (who haven't tested our extension) and 100% of the Group 2 (who have tested our extension) preferred the modified version of Introduction page (Version 2), which is adding an "About" page to the interface. In total, 54.5% participants preferred the modified version.

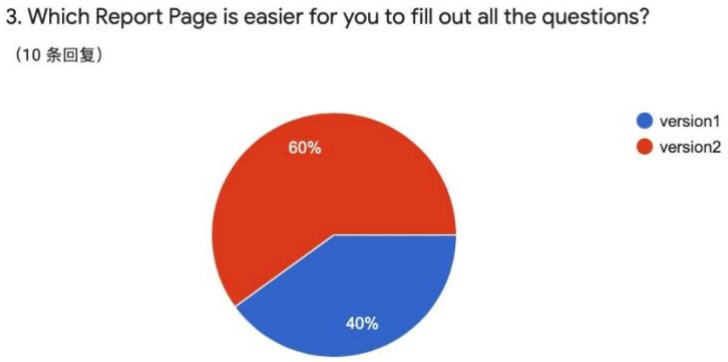


2. Among all the responses, none of the users in Group 1 (who haven't tested our extension) and 100% of the Group 2 (who have tested our extension) think the modified version of UI (Version 2) provides more clear and detailed information about the detected Dark Patterns. In total, 45.5% responses preferred the modified version.

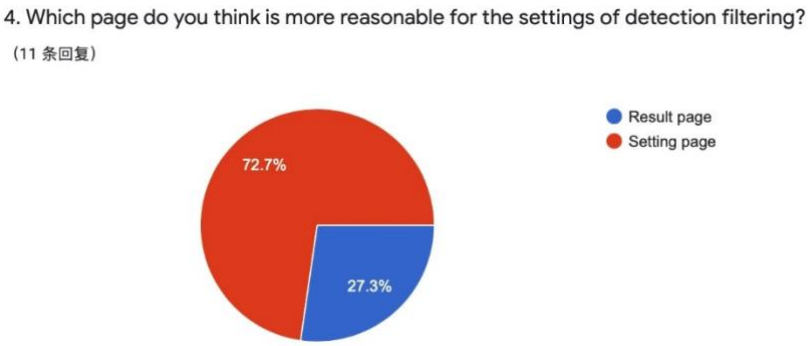


3. Among all the responses, 33.3% of the Group 1 (who haven't tested our extension) and 100% of the Group 2 (who have tested our extension) preferred

the modified version of Report page (Version 2). In total, 60% participants preferred the modified version.



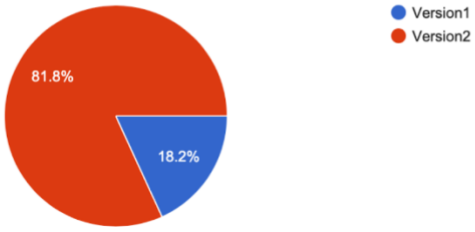
4. Among all the responses, 50% of the Group 1 (who haven't tested our extension) and 100% of the Group 2 (who have tested our extension) preferred the modified version (Version 2), which is to put the filter in the settings page. In total, 72.7% participants preferred the modified version.



5. Among all the responses, 66.7% of the Group 1 (who haven't tested our extension) and 100% of the Group 2 (who have tested our extension) preferred the modified version of the highlighting icons(Version 2). In total,81.8% participants preferred the modified version.

5. Which Highlight style is more eye-catching?

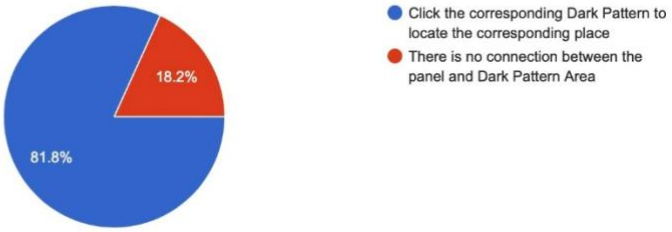
（11 条回复）



6. Among all the responses, 66.7% of the Group 1 (who haven't tested our extension) and 100% of the Group 2 (who have tested our extension) thought the revised version was better. Overall, 81.8% thought we should build stronger links between panels and Highlighting areas.

6. Which version of UI provides a closer connection between the results on the UI and the highlighting on the web page?

（11 条回复）



4.3 Results of User Evaluation with Extension Version 2

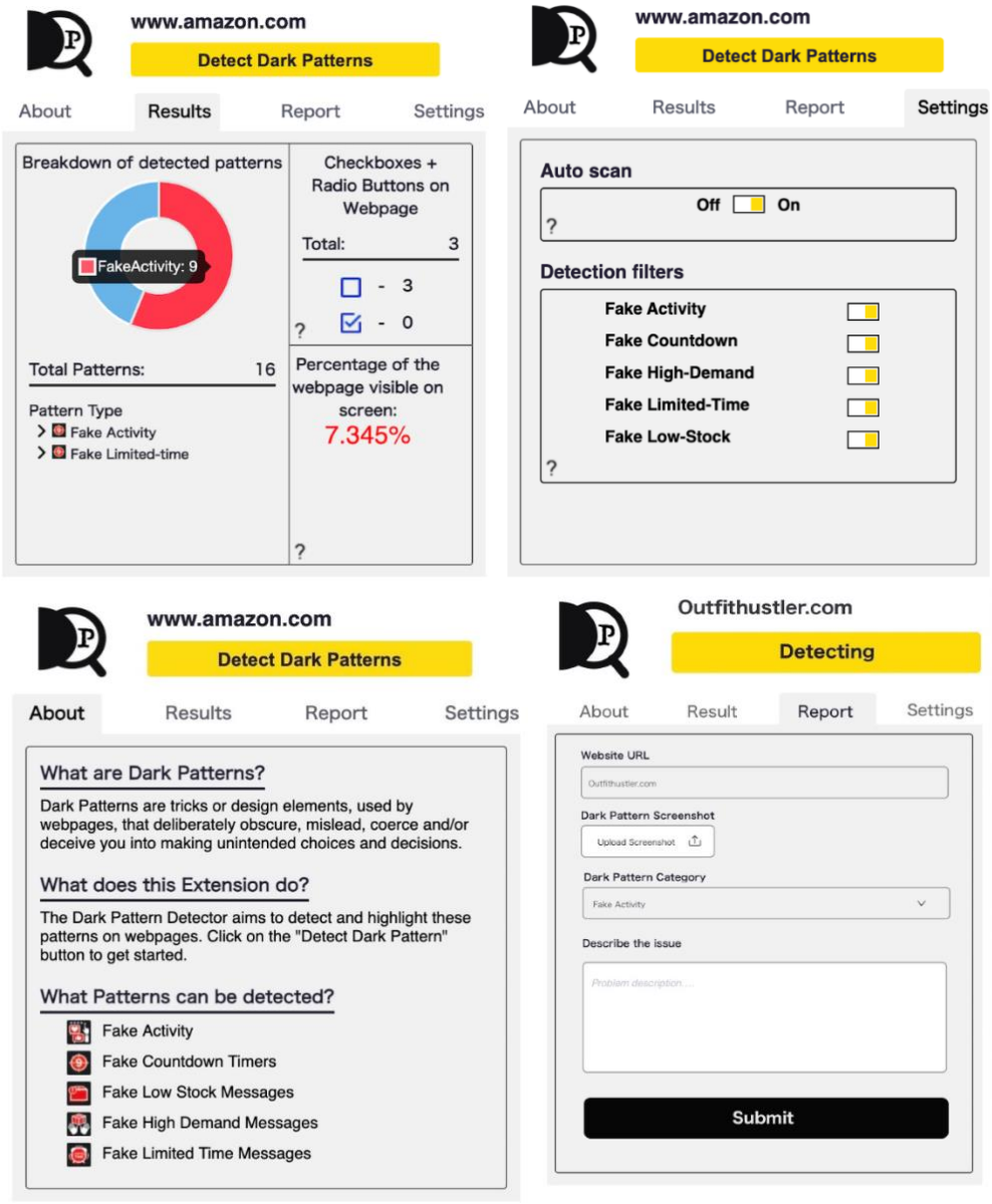


Figure 2 - Layout of Extension Version 2

4.3.1 Cognitive Walkthrough

For the Cognitive Walkthrough for extension version 2, 8 tasks were conducted by users. The tasks are shown as below:

1. How would you find more information on Dark Patterns?
2. How would you detect Dark patterns on a webpage?
3. How would you report a dark pattern on a webpage?
4. How do you turn on/off Autoscan?
5. How do you turn on/off the detection of certain types of patterns?
6. How would you locate highlighted patterns on webpages?

7. Can you find what each icon represents?
8. Where do you find information about checkboxes and page location, in the extension?

	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7	Task 8
	Task Completion / Time Taken							
User 1	F/S	S/S	S/S	S/S	F/F	F/S	F/F	S/F
User 2	F/S	S/S	S/S	S/S	F/S	F/F	F/S	S/S
User 3	F/S	S/S	F/F	S/S	S/S	F/F	S/S	S/S
User 4	F/S	S/S	F/F	S/S	S/S	S/S	S/S	S/S
User 5	F/S	S/S	F/F	S/S	S/S	S/S	S/S	S/S
Task Score	0%	100%	40%	100%	60%	40%	60%	100%
Time Score	100%	100%	40%	100%	80%	60%	80%	80%

Table 10 - Performance results obtained through Cognitive Walkthrough

From the result of the tasks, we could see at least 3 functionalities need to be improved, the issues are as below:

1. Most of the users don't have the patience to read the text description of dark patterns on the 'About' page of the extension. (Task 1)
2. The "Report" function on the extension, again, users found it difficult to understand what should be filled in in each field as in Version 1, and it took users quite long to fill all the fields and submit the report. (Task 3)
3. Users don't know how to locate the highlighted dark patterns according to the results on the extension 'Result' page. (Task 6)

Performance results: 80 responses in total (40 Task Completion Status responses + 40 Time Taken responses) are organised according to the measuring criteria. Among all the responses, 57 were successful, and 23 were failed. All the success points contribute to the success rate and all the failure points contribute to error rate.

The success rate of Cognitive Walkthrough for Version 2 would be: 57/80 = 71.3%. The error rate of Cognitive Walkthrough would be: 1- 71.3% = 28.7%.

4.3.2 Questionnaire (Quantitative)

For the questionnaire, 9 quantitative questions are presented to users, and 5 metric parameters are used for result evaluation, the results are shown in Table 11.

	Readability	Data Relevance	Accessibility	Consistency	Visual Weight
User 1	S(4)	F(3)	S(4)	F(3.7)	F(2)
User 2	S(5)	S(4.5)	S(4)	S(4)	S(4)
User 3	S(5)	S(5)	S(4)	F(3.3)	S(5)
User 4	S(5)	S(4.5)	S(5)	S(5)	F(3.5)
User 5	S(4)	F(3)	S(4)	F(3.3)	F(3.5)
Average Score	4.6	4	4.2	3.9	3.6
Success Score	100%	60%	100%	40%	40%

Table 11 - Results obtained from questionnaire (quantitative)

From the result of the questionnaire, some of the designs still need to be adjusted to improve the usability:

1. Consistency needs to be improved, for example, when turning on the auto-scan, how to tell the user that our extension is still working.
2. In terms of Visual Weight, the pie chart was not very attractive for the users, however it occupies more than ¼ of the result page.

Questionnaire results: 25 responses in total are organised according to the measuring criteria. Among all the responses, 17 were successful, and 8 were failed. All the success points contribute to the success rate and all the failure points contribute to error rate.

The success rate of Cognitive Walkthrough for Version 2 would be: 17/25 = 68%. The error rate of Cognitive Walkthrough would be: 1- 68% = 32%.

The **final success rate** is the average of the success rate of Cognitive Walkthrough and the success rate of Questionnaire, which would be (71.3% + 68%) / 2 = 69.7%, and the final error rate is 1 - success rate = 30.3%.

4.3.3 Think Aloud

Data (in Table 12) included in the Think Aloud are:

1. Qualitative results from the questionnaire
2. Issues the users meet during the Cognitive Walkthrough
3. Feedback and suggestion users gave during and after the Cognitive Walkthrough

	Issues
Functionality	Detection process took too long without any timer, the user thought it crashed, there should be a timer.
	User doesn't know the result is clickable to be directed to the dark pattern location on the page.
	Don't know what each dark pattern means in the "settings" page, suggest a hover effect on each of them with a picture pop-up for explanation.
	Should have a "help/FAQ" function regarding how to use the extension.
	Perhaps a way to contact developers to report issues and or bugs.
	Can every webpage be detected automatically? Or just one website page
	Connect the Dark Pattern Introduction Website with the content on the "About" page.
Layout	The pie graph is not wanted by the user, which occupies the most space in the result page.
	Move the About page back or change it to the scroll down menu
	Logos of the dark patterns are too small and need to be clearer in the extension
	Layout of reports page and results page need to be better
	Bigger icons for dark pattern results, explanation it
	Change the icon colour or its style, the red and black colour icon makes the user anxious.
	Result should be at the first page, don't want to see "About" every time when initiating the extension
	Thought the 'Report' tab meant a report of the detection. i.e., 'Results'

	Did not realise there was a drop-down menu under each detected type of pattern
	Move the link for DP detector website further up the page it is on. Change 'report' to 'report dark pattern'.
	The fact that if there is no check boxes on the page, that section of the "Results" tab is blank, and it doesn't say "No buttons are detected on this page".
	I like the size of the extension, but I would like if it had a control bar on the top with an “X” on the top right to close the window, and in the middle of control bar, a piece of text saying “[A][A][A]” which could make the window (and text and diagrams) bigger, and a lot bigger.
	The layout on reports page and results page needs to be better
Content	Bigger fonts needed in the extension (for detection results), or replaced with image/gif when hovering on the text
	The “?” should be beside each heading, not the left corner of the box
	No idea what “Detection Filter” means in the “settings” page
	Change 'report' to 'report dark pattern'
	Colour needs to be unified
	Don't know what the words below pattern type stand for, so better to mark them directly, or tell the customer the truth
	Report page questions use need to apply better explanations on what to fill out in these sections
	Why not just write about rational consumption? Instead of just using icon
	Highlight area the user wants something similar to iPhone/Android warning window, a grey exclamation mark inside a yellow triangle sign.

Table 12 - Results obtained from Think Aloud

4.3.4 Expert Users Review

Severity Ratings	
Rating	Definition

0	NO PROBLEM
1	Cosmetic Problem
2	Minor Usability Problem
3	Major Usability Problem
4	Usability Catastrophe

Table 13 - Nielsen's Severity Rating

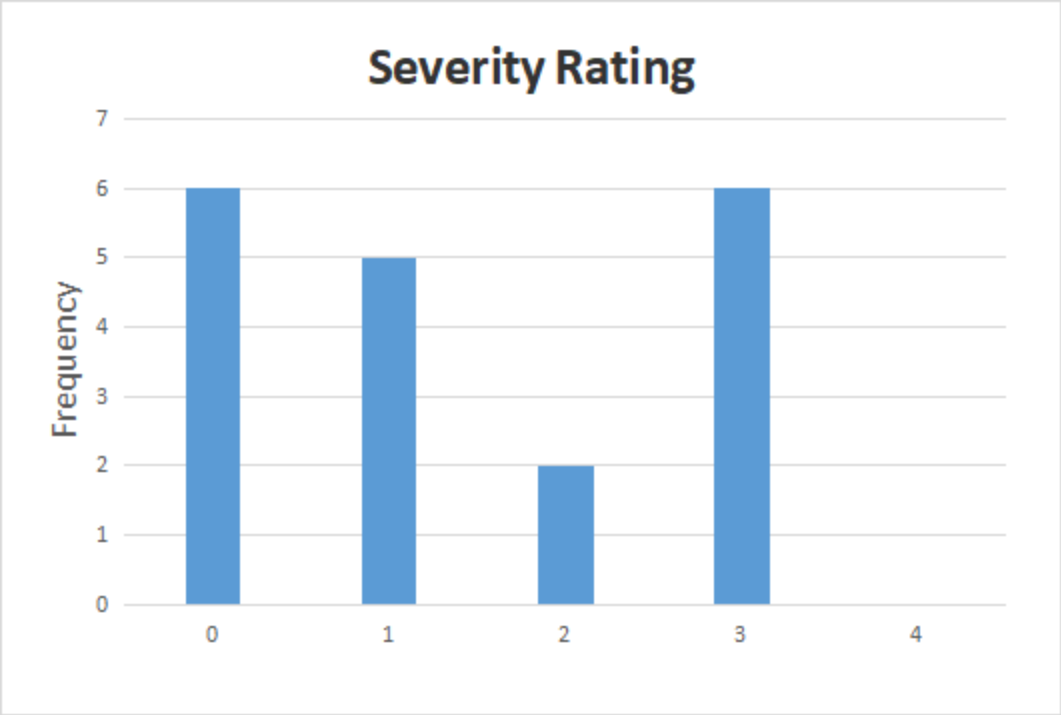


Figure 3 - Frequency Severity Rating in Dark Pattern Extension

As can be seen from the figure above, the most serious problems are at the severity level (3), with a total of six . They respectively violate Nielsen heuristics usability principles of "Visibility of system status", "Recognition rather than recall", "Flexibility and efficiency of use", "Help and documentation", these four principles. In addition, severity level (2) has two problems, which belong to the "Match between system and the real world (real user's habit)" and "User control and freedom". There are also five problems which are Cosmetic Problems.

Severity Ratings	Nielsen heuristics	Problems
3	Visibility of system status	1.Once auto-detect is on, patterns are being detected even when going on to new pages? There is nothing indicating this.
		2.Difficulty to see what the icons are.

		3.Difficult to see where patterns are actually on the page
	Flexibility and efficiency of use	4.The idea of being able to see exactly where dark patterns are is a good one but not accurate at the moment.
	Recognition rather than recall	5.Provides me with no more information on dark patterns. Need to connect to the information website.
	Help and documentation	6.Website needs to be incorporated to educate the user.
2	Match between system and the real world (real user's habit)	Slider buttons are good. Should have labels on them.
	User control and freedom.	The ">" symbol at the start of each pattern isn't obviously a menu option to see more detail, maybe change to [>] ??
1	Consistency and standards	Is the website consistent with the add-on?
	Match between system and the real world (real user's habit)	Extension defaults to the "About" tab, it gets boring after a while.
	Flexibility and efficiency of use	I'd like if I could move around elements on the "Results" tab, or replace the Pie Chart with a Bar Chart
	Aesthetic and minimalist design	Lots of blank space in the "Report" tab
	Help and documentation	Good "About" tab, but I would like a 50 second video showing the key features of the extension

Table 14 - What each problem is at each severity level

.

5 Conclusions

After running this evaluation study, we can conclude that the solutions that were implemented after evaluation phase 1 were sufficient in most instances but not all. For example, in the first phase of the cognitive walkthrough, the data relevancy and over design of the UI was an issue. The solution to this problem was to revamp the UI design as seen in figure 2, which include more relevant data and a more eye-catching design. It can be concluded that these solutions were of a high enough quality to solve the problem as in Phase 2, users did not believe that the data relevancy and eye-catching design was an issue. But, due to the revamping of the UI, users found other issues, like the amount of information on the “About” page and “Report” page confusion was a recurring issue from Phase 1. This shows that the quality of the solution implemented after the evaluation of phase 1, concerning the “Report” page, was not of a good enough standard to completely solve the issue.

In a future evaluation study, a more variety of users could be obtained as a lot of our average users were around the same age group, with not a lot of representation from the older average user. During the first evaluation of the system, we only got the average users to do the evaluations, it would have been more helpful to have gotten some expert users to do evaluations as well and throughout the entire design process.

If another evaluation was to occur, having more open-ended questions would be more beneficial, as they provided us with better insights to what was wrong with extension compared to questions based on the Likert Scale, even though the Likert Scale questions were very informative about the how the users felt about the extension.

Overall, the user evaluation gave us some very important insights to what the users liked and disliked about the extension, as well as what tasks they struggled to complete based off of our design. A lot of these design flaws were things that we might have overlooked or just assumed that the user would understand something else simply because we understood it. However, this is not the case a lot of the time and from the results gotten, there are some major design flaws that need to be addressed.

References

Co-discovery Learning. (n.d.). Understanding Chi. Retrieved November 17, 2021, from http://hci.ilikecake.ie/eval_codiscoverylearning.htm

Dalrymple, B. (2018). *Cognitive Walkthroughs*. Medium. Retrieved 17 November 2021, from https://medium.com/user-research/cognitive-walkthroughs-b84c4f0a14d4.

Dewi, P. W. S., Dantes, G. R., & Indrawan, G. (2020). User experience evaluation of e-report application using cognitive walkthrough (cw), heuristic evaluation (he) and userexperience questionnaire (ueq). *Journal of Physics: Conference Series, 1516,* 012024. https://doi.org/10.1088/1742-6596/1516/1/012024

Easy Checks – A First Review of Web Accessibility. (n.d.). W3C. Retrieved November 15, 2021, from https://www.w3.org/WAI/test-evaluate/preliminary/

Hes, B. (2017). *Data relevance – Crucial in Asset Performance Management*. Stork. Retrieved 17 November 2021, from https://www.stork.com/en/about-us/blog/data-relevance-crucial-in-asset-performance-management.

IshΔn. (2019, February 3). *What and Why of Usability Evaluation*. Medium.https://blog.prototypr.io/what-and-why-of-usability-evaluation-46bf4b6dee07

Kuar, A. (2018). *Accessibility guidelines for UX Designers*. Medium. Retrieved 17 November 2021, from https://uxdesign.cc/accessibility-guidelines-for-a-ux-designer-c3ba775539be.

Mcleod, S. (n.d.). *Likert Scale Definition, Examples and Analysis | Simply Psychology. Simply Psychology.* Retrieved November 15, 2021, from https://www.simplypsychology.org/likert-scale.html

NI Business Info. (n.d.). *Consistency in web design | nibusinessinfo.co.uk*. Retrieved November 15, 2021, from https://www.nibusinessinfo.co.uk/content/consistency-web-design

Sauro, J., PhD. (2013, July 30). *Rating the Severity of Usability Problems – MeasuringU.* Measuring U. Retrieved November 15, 2021, from https://measuringu.com/rating-severity/

What is Aesthetics?. (2021). Retrieved 17 November 2021, from https://www.interaction-design.org/literature/topics/aesthetics

Xiangyu Wang. (2008). Using Cognitive Walkthrough procedure to prototype and evaluate dynamic menu interfaces: A design improvement. *2008 12th International*

Conference on Computer Supported Cooperative Work in Design. Published. https://doi.org/10.1109/cscwd.2008.4536959

Appendix

1. AB Test Survey
   https://docs.google.com/forms/d/e/1FAIpQLSeixQCIZsumVlAk7wJzBkjshm-P8ynHCwZAOcW8alv4eNh_jA/viewform?usp=sf_link

2. Questionnaire after cognitive Walkthrough
   https://docs.google.com/forms/d/e/1FAIpQLSeH3KwozqdBrn_iFb7BJHh8d_v1gtqpCKunxICG_JzewC4tYg/viewform?usp=sf_link