Analysis of Item Presentation and Search Time in Online Shopping

Walter W. Thompson Jr.
Clemson University
240F Campus Drive, Clemson, SC 29630
waltert@g.clemson.edu

Abstract

A foundational practice in commercial retail is to present items to customers in the most efficient way possible. Companies have always strived to present their goods and services in the most efficient way that they can. Today, you can see the result of countless consumer studies if you walk into Walmart, Target, or The Home Depot. Virtually every item that you see from the moment that you walk into the door has a strategic purpose. In recent years, online shopping has begun to dominate the market. The rise of online shopping has pushed more and more companies to embrace online retail in order to compete. Even smaller companies now have a way to reach consumers without spending money on brick and mortar stores. This intense competition has bought about a new question: What is the best way to present goods to consumers online? In this study we use eye tracking methodologies to examine the effects of changes in the presentation of search results for online shopping. We examine the two most common search result item presentations —a grid and a list —and analyze them to find if one is more efficient than the other. **(Abstract will be refined and more information will be added after the study is complete and the results are known)**

Keywords—Online Shopping; Grid; List; Item Presentation; Target Item

Introduction

As online shopping has grown larger, the need for each company to be more efficient and unique has grown exponentially. Many companies use search engines to allow consumers to find the items that they are looking for. Search engines allow consumers to merely type a simple word or phrase and be greeted with an entire screen of items that best matched their search. Most search engines provide results that are ordered from most to least relevant. In online shopping, results usually include a picture of the item, the item's price, and some lesser data such as the product number. Most of these search engine results are quite efficient, but it is important to know the ways in which people view these results. Most online shopping websites provide a tool that allows consumers to customize the presentation of their search results. The two most common customization options are to present the results as a list or as a grid. A list is a single column of items that can have as many as fifty to one-hundred rows of results on a single page. The grid option can vary in the number of columns, but it usually has three or four. A grid can have many rows as a list; however, some websites will reduce the number of rows per page so that the number of results per page will be the same as a list. Examining these two item presentations will likely lead one to ask which one is "better". This study asks participants to complete a string of visual search tasks while using an eye tracker to follow their gaze. The goal of this paper is to provide an in-depth analysis on the differences of how consumers view a list and a grid presentation. Our second goal is to determine which one is actually more efficient in providing users with the results that they are looking for. Efficiency can be examined by analyzing the time a user takes to find the item that they are looking for. The Home Depot's website will be used as the online store for the experiment. The two searches that are used to analyze the list and the grid are "drill" and "hammer". We hypothesize that the list will be more efficient than the grid in terms of finding an item in the shortest amount of time possible. Our general reasoning for this is based partially on the work of Qiuzhen Wang in his 2014 study that is discussed further in the next section. In his study, he found that more information slowed users by distracting them. We believe that users will almost certainly traverse the list from top to bottom, but the grid may introduce some hesitation. We do, however, expect the grid to have a much wider variation of results due to all the different ways that it can be traversed. The item locations within the list will likely remain within a relatively predictable range of search times. In addition, we predict the "hammer" searches to produce relatively faster search results due to having more simplistic names and appearance.

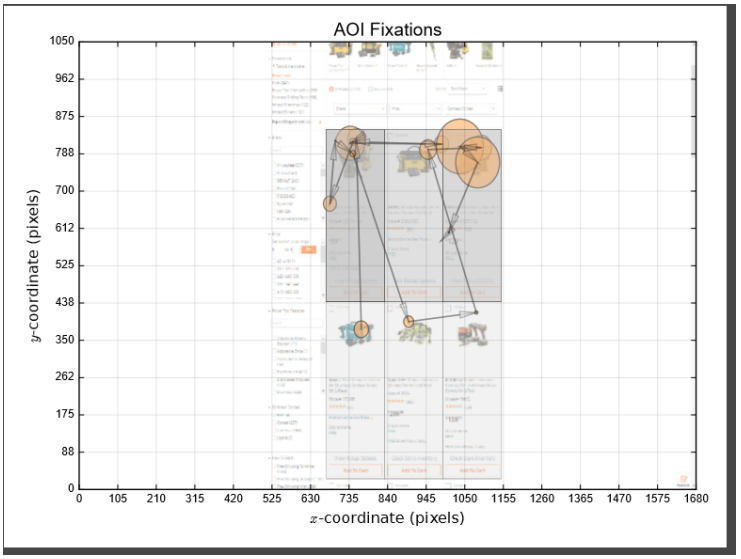

Figure 1. Visualization of a participant's search for a target item on a grid presentation. The colored boxes represent areas of interest.

Background

Using eye tracking methodology to determine the effectiveness of item presentation is nothing new. In fact, it's been invaluable for many online and physical retailers. Eye tracking has been used to examine how consumers view shelves and select brands at physical retail stores. A fairly well know example of this is the work done by P. Chandon, J. W. Hutchinson, and S. H. Young. In their 2002 study entitled, "Unseen is Unsold: Assessing Visual Equity with Commercial Eye-Tracking Data" they found that brand selection is very difficult to fully account for. Consumers would often walk into the store with previous thoughts

about particular brands. While there are a number of eye tracking studies that examine physical retail stores and brand selection, examining online retail websites is traditionally easier. Physical retail often requires a more intrusive apparatus that could lead to more unnatural shopping experiences. On the other hand, online retail eye tracking studies can use stationary non-intrusive eye tracking devices like the Gazepoint GP3 used in this paper's study. Study participants who can operate without an intrusive eye tracking device are less likely to give abnormal results in a study and will be able to search for products more naturally.

In many ways, one of the roots of online shopping is the search engine. It is, after all, the key to finding the item that you want to buy. A complete understanding of search engines is needed before a search engine study can be accurately conducted. To fully understand search engine results, one must study the goal of the search. A 2005 study by Uichin Lee, Zhenyu Liu, and Junghoo Cho examines features that can be used to identify user search goals. Their study is entitled, "Automatic identification of user goals in web search" and it provides a better understanding of some of the difficulties that search results can encounter when attempting to meet the goals of a user. Aside from the search engines, the design decisions of a company can have an effect on how efficient a search is. An interesting study that examines this problem in online shopping was performed by Qiuzhen Wang in 2014. His study is titled, "An eye-tracking study of website complexity from cognitive load perspective". He found that when participants performed simple navigation tasks, task completion time on the websites with high complexity and medium complexity was higher than that on the website with low complexity. This means that providing too much information is a real issue for website creators. Many companies want to provide as much user information as possible to ensure that users have extensive information about their goods and services. However, this study found that users' attention is easily distracted as website complexity increases. This study shows the benefits of retailers having a clean, efficient, and streamlined website. Keeping these facts in mind, it is important to consider the complexity of an online shopping site when conduction an experiment. This is in part why The Home Depot's website was selected for the study in this paper. The website is very streamlined and has very few distractions for the participants.

While studies such as these have led to the work in this paper, this experiment is directly inspired by the work of Edward Cutrell and Zhiwei Guan. Their 2007 study entitled, "What Are You Looking For? An Eye-tracking Study of Information Usage in Web Search" built the groundwork for the study in this paper. In many ways, this paper could be considered a continuation of their study. In their study, they examined the effect of changes in presentation for web search results. Their study involved giving navigational and informational tasks to participants and analyzing the results. Navigational tasks are that same type of tasks that we use in our paper to test the efficiency of the list item presentation vs. the grid item presentation.

Methodology

Apparatus

For this study, a Gazepoint GP3 standalone eye tracker was used. The eye tracker has an accuracy of half a degree, about 50 pixels. The Gazepoint runs at a sampling rate of 60Hz and is mounted at the bottom of a computer monitor. The monitor used for the experiment was a 22" Dell screen and was set at a resolution of 1680 x 1050. The GP3 eye tracker is very non-intrusive and does not have to physically make any contact with the participant in order to accurately track eye movement. The only setup that is required is a short 9-point calibration test. The tracker does not distract the user by using flashing lights or emitting distracting noises. This allows the experiment to be as representative and accurate as possible.

Participants

The goal of the study in terms of participation was to have 16 participants in order to have four participants run through each of the four possible stimuli orderings. A total of 19 participants took part in this study. Their ages ranged from 19 to 50 with the median age being 21.5. All of the participants were either Clemson University students or professors. Of those participants, 16 of them were male and 3 were female. Three participants' data had to be removed from the study due to calibration issues and/or incorrectly answering a large number of search tasks. This resulted in the final number of participants being 16 with 14 of those being male and 2 being female. All participants were recruited using word of mouth.

Stimulus

The stimuli for this experiment includes a visual search task. A website needed to be selected as the primary images for this task to be performed. We have selected The Home Depot's website due to its efficient and non-distracting interface. Throughout the experiment's search tasks, participants will see images of Home Depot search results and have to find certain items in those results. There are two types of searches and two types of item presentations. The first type of search is for a drill. The participant is given the "drill" search results in both a list and a grid format. The items in the lists and grid formats are the exact same; they are merely in different presentations. The two "drill" search images that were shown to participants can be seen in Figure 1 and Figure 2. The second type of search is for a hammer. The "hammer" search results also appear in the list and grid format. Images for the "hammer" search results can be found in Figure 3 and Figure 4. Areas of interest(AOIs) were generated around each individual target item to track the number of fixations that each item receives.

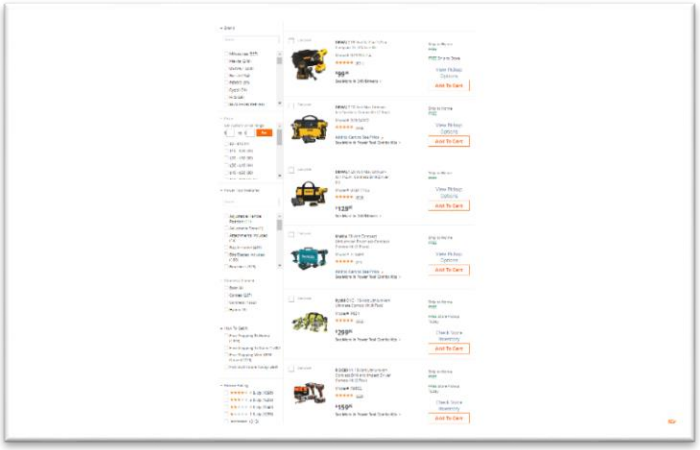

Figure 2. The drill item type in a list presentation. One of the four different visual stimuli for the search task.

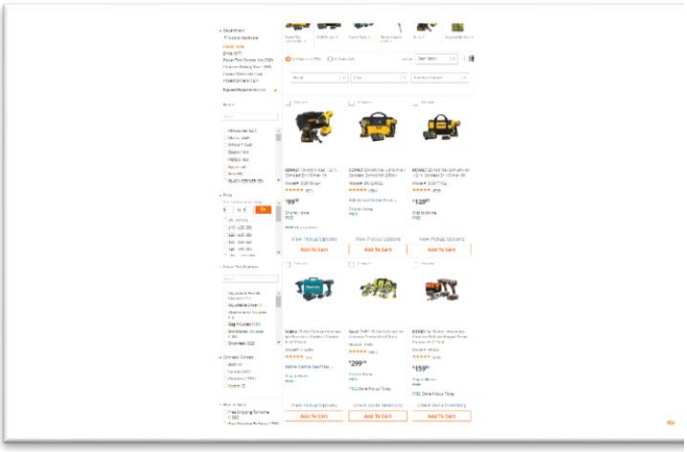

Figure 3. The drill item type in a grid presentation. One of the four different visual stimuli for the search task.

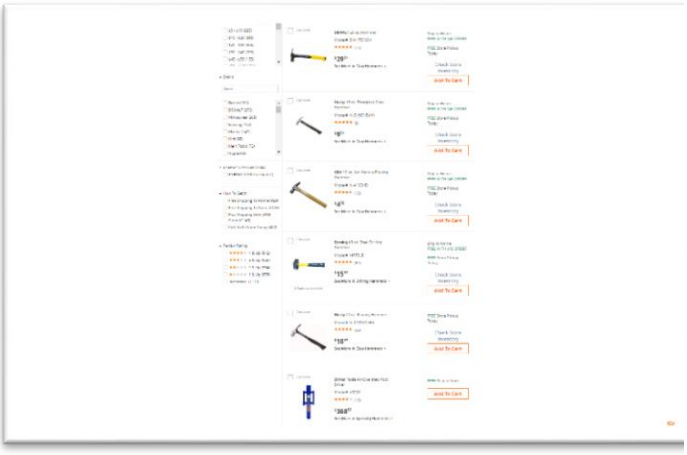

Figure 4. The hammer item type in a list presentation. One of the four different visual stimuli for the search task

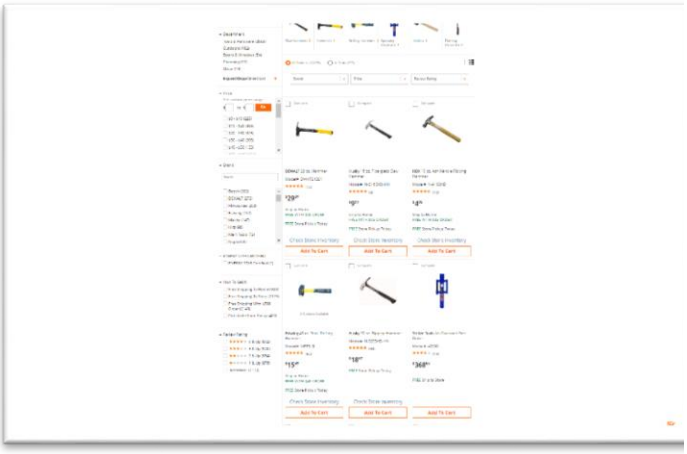

Figure 5. The hammer item type in a grid presentation. One of the four different visual stimuli for the search task.

Procedure

Before the experiment begins, participants are thoroughly informed about the nature of the task they will be performing. They are informed that they will be viewing images of online shopping search results and will have to find a target item within those search results. Each participant must then take a short 9-point calibration test to set the eye tracker. Once the calibration test is complete, participants are subjected to a brief visual search task to practice and ensure that he or she understands the task. The practice search task begins when the participant is shown a grey screen with a written prompt to find a Dewalt scroll saw. An image for "scroll saw" search results then appears. The participant had to find the Dewalt scroll saw in the list presentation and then the grid presentation. These results are not recorded or considered important in any way. They are simply used for teaching the participant how the visual search task is performed. Once the participant finishes the practice problem, he or she is given an opportunity to ask any further questions. When the participant is ready, the name of the item that the participant must find appears onscreen. The name of the target item is also read aloud to them by the experimenter. Similar to the practice task, the participant is be given a target item and shown an image search. This time, they are shown results for a "drill". The participant has to perform a visual search task to find the requested drill in the list and the grid presentations. The previous steps are then repeated with the search results for "hammer". To see the orderings in which the task types are given to participants, see Table 2.

Table 1. An example of the target items that are received by participants. Half of the tasks in each category are in the list presentation and the other half are in the grid presentation.

Drill
• **Find a Dewalt 18-Volt Ni-Cad Compact Drill** • **Find a Dewalt 20-Volt Cordless 2 Tool Combo Kit** • **Find a Dewalt 20-Volt Max Cordless Drill** • **Find a Makita 18-Volt Cordless Combo Kit** • **Find a Ryobi ONE+ Ultimate 6 Tool Combo Kit** • **Find a RIGID X4 Cordless Drill**
Hammer
• **Find Dewalt 20 oz. Hammer** • **Find Husky 16 oz. Fiberglass Hammer** • **Find HDX 10 oz. Ash Handle Ripping Hammer** • **Find Estwing 48 oz. Steel Drilling Hammer** • **Find Husky 20 oz. Ripping Hammer** • **Find Striker Tools Air-Operated Post Driver**

Table 2. The different combinations of search task types that a participant can receive. Each task includes finding 3 target items.

A=grid, hammer
B=grid, drill
C=list, hammer
D=list, drill

Participant	Task 1	Task 2	Task 3	Task 4
P1	A	B	C	D
P2	B	C	D	A
P3	C	D	A	B
P4	D	A	B	C

Experimental Design

The experiment crosses Item Presentation (2) x Item Type (2). In these factors, item presentation includes the list and grid presentations. The item type factors include the drill and the hammer searches. Both factors in the experiment are within subject. In the tasks, the participant's only goal is to reach the target item. The target items for every participant were the same, but the order in which the tasks were given was randomized for each participant. Every target item task can be seen in Table 1. There are a total of four different orderings for Item Presentation (2) x Item Type (2) that a participant can receive. The different combinations can be viewed in Table 2. Each task in Table 2 represents a total of three target items that must be found. This means that there is a total of twelve search tasks for every participant. In addition to the counterbalance put in place to ensure that item presentation is evenly distributed, the order in which target items are assigned was randomly generated to form a unique pattern within each group of four participants.

Results

For analyzing the data in this study, we examined the gaze fixations that landed within our defined areas of interest. In order for a participant's gaze to be considered a fixation, his or her gaze must have landed on a point in the area of interest for at least 100 ms. The defined areas of interest for the grid can be seen as the colored rectangles in Figure 1 and the areas of interest for the list can be seen in Figure 6 and Figure 7. In order to determine if one of the item presentations is more effective than the other, we looked at two major factors. The first factor was how much time had elapsed by the time that the participant ended the search by hitting the spacebar. This meant that he or she had reportedly found the item. The second factor that we looked at was the number of non-target items that were viewed before the participant hit the spacebar.

General analysis of search results

Generally speaking, the color of the item played a large role. In the drill search tasks, there were three similar yellow Dewalt drills. Items such as the Dewalt 20-Volt Cordless 2 Tool Combo Kit had noticeably higher search times than items with more standout colors, such as the red Rigid X4 drill. In addition to longer search times, items with a standout color and/or appearance would often have very basic and uncomplicated scan paths when compared to target items with similar colors. Looking at Figure 7, we can see a participant searching for a 20 volt cordless DeWalt drill. Notice that the participant's gaze is repeatedly going back and forth fixating on the three yellow DeWalt drills. Comparing Figure 6 with Figure 7, we can see a very different result. The participant's gaze is very linear, starting at the upper part of the list and then quickly jumping to the target item. To put both Figures into perspective, the search in Figure 6 lasted approximately 5.7 seconds and the search in Figure 7 lasted approximately 3.2 seconds. These kinds of patterns in regards to color were seen across all item presentations regardless of whether or not the participant was shown items in a list or a grid format. We wanted to analyze how significant this difference was. Figure 8 shows the mean time to find each of the three yellow DeWalt drills versus the mean time to find each of the three other drills on the search tasks.

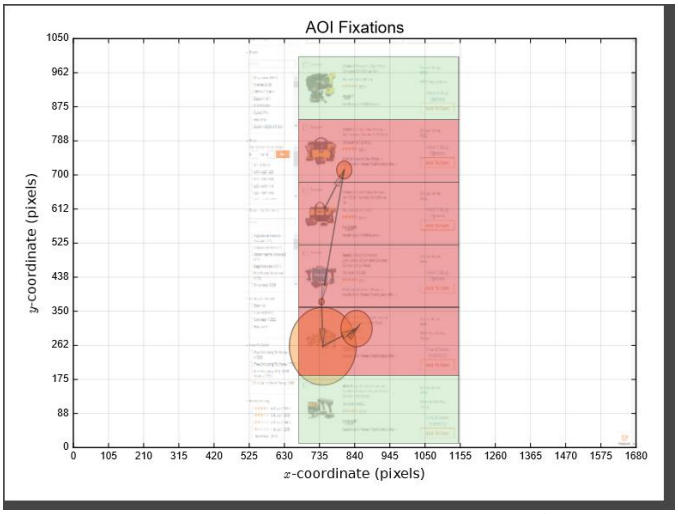

Figure 6. This common short scan path shows the gaze of a participant searching for the Ryobi ONE+ drill on a list presentation. The Ryobi drill's bright green color and unique picture help it to stand out from the other results. These factors often led to very fast search times and few fixations before finding the desired item.

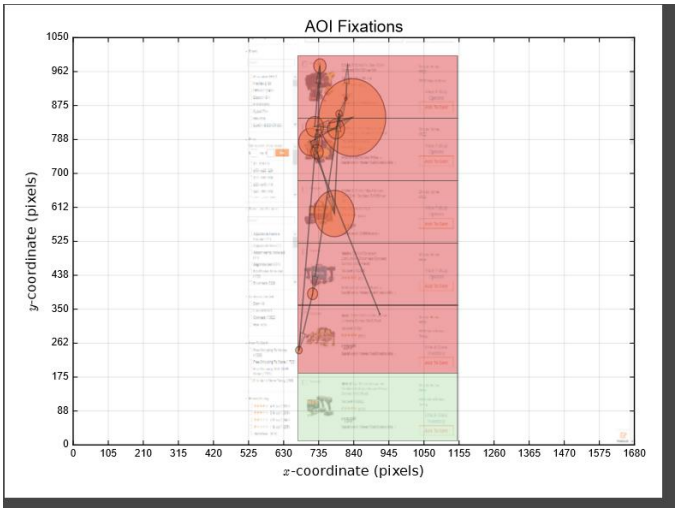

Figure 7. This common long scan path shows the gaze of a participant searching for a 20 volt DeWalt cordless drill combo on a list presentation. The Dewalt drill's appearance is quite similar to the two other Dewalt drill items on the list. This often led participants to thoroughly traverse the search results looking for the target item.

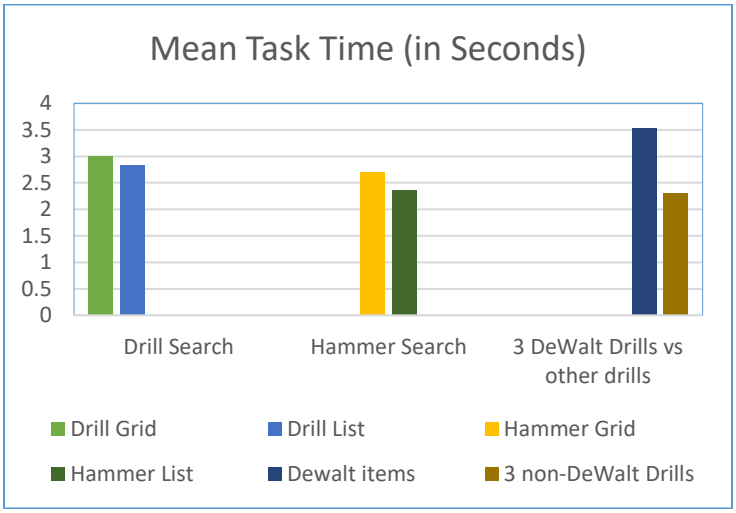

Figure 8. Visualization of the mean task completion time for Drill grid tasks was 3.01 seconds. The Drill list mean was 2.84. The Hammer grid and Hammer list mean completion times were 2.70 and 2.37 respectively.

Wilcoxon Scores (Rank Sums) for Variable time Classified by Variable presentation					
presentation	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
grid	48	2286.0	2328.0	136.318358	47.6250
list	48	2370.0	2328.0	136.318358	49.3750
Average scores were used for ties.					

Table 3. A Rank Sums test with the goal of determining if there is a statistically significant difference between the underlying distributions of search time for the Drill List and the Drill Grid.

To see if there was a difference in the distribution of search task time between the list and the grid, we ran a Ranked Sums test on the Drill search results. Tables 3 and 4 show the results of the test. The test indicated that there was insufficient evidence to conclude that the distribution of time in the list and grid presentations are significantly different (z=-0.3044, p=.7608). This was not particularly surprising after we found the mean search times for the drill search to be very similar. Then we examined the hammer search times. The mean values indicated a slightly larger and potentially more promising difference between the hammer list and the hammer grid. However, despite the hammer list and grid having a larger variation between the two means, the median was not enough to be statistically significant (z=1.0748, p=.2825). The results in regards to time elapsed while searching for the target item stayed relatively flat. This seems to indicate that the participants were not significantly hindered by the list or the grid presentation.

We examined the number of non-target items that each participant fixated on before they hit the spacebar. Hitting the spacebar, as described earlier, is the participant's tool to convey that they are confident that the item has been correctly identified. In this analysis, non-target items were included regardless of whether or not the item was viewed before or after the target item. If an item was viewed after the participant found the target item, this was seen as the participant confirming that they had the correct result by quickly examining a nearby item. After collecting the total number of non-target items viewed, we found that each of them have fairly similar results that can be seen on table 7. We are not entirely certain why the number of non-target fixations are lower for the Hammer Grid. This is rather odd considering that the Hammer List has the most non-target fixations. We considered that the Husky hammers with their identical pictures could somehow be involved. However, the Hammer List presentation includes all of the same target items as the Hammer Grid.

Drill Grid	Drill List	Hammer Grid	Hammer List
81	87	63	89

Table 7. The total number of non-target items that participants fixated on.

Wilcoxon Two-Sample Test	
Statistic	2286.0000
Normal Approximation	
Z	-0.3044
One-Sided Pr < Z	0.3804
Two-Sided Pr > |Z|	0.7608
t Approximation	
One-Sided Pr < Z	0.3807
Two-Sided Pr > |Z|	0.7615
Z includes a continuity correction of 0.5.	

Table 4. Results of the Rank Sums test for the "Drill" search. The Results indicate that there insufficient evidence to conclude that there is a statistically significant difference between the time distributions of Drill grid and Drill list (z=-0.3044, p=.7608)

Wilcoxon Scores (Rank Sums) for Variable time Classified by Variable presentation					
presentation	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
gridh	48	2475.0	2328.0	136.304921	51.56250
listh	48	2181.0	2328.0	136.304921	45.43750
Average scores were used for ties.					

Table 5. A Rank Sums test with the goal of determining if there is a statistically significant difference between the underlying distributions of search time for the Hammer List and the Hammer Grid.

Wilcoxon Two-Sample Test	
Statistic	2475.0000
Normal Approximation	
Z	1.0748
One-Sided Pr > Z	0.1412
Two-Sided Pr > |Z|	0.2825
t Approximation	
One-Sided Pr > Z	0.1426
Two-Sided Pr > |Z|	0.2852
Z includes a continuity correction of 0.5.	

Table 6. Results of the Rank Sums test for the "Hammer" search. The Results indicate that there insufficient evidence to conclude that there is a statistically significant difference between the time distributions of Drill grid and Drill list (z=-1.0748, p=.2825)

Discussion

Unfortunately, the results of the study seem to tell more of a story about the differences between target items with standout colors and those without. Based on the results of the study, it appears that items with standout colors resulted in an easy image matching task for the participants. Very similar items, like the three DeWalt drills or the Husky hammers, caused the participants much more trouble. A few participants even hit the spacebar when they had the wrong item.

This was actually a recurring problem throughout the experiment. Some participants' results had to be removed from the study due to "rushing" the experiment and giving many wrong answers. This problem persisted despite the fact that each participant was informed that accuracy should be valued much more than speed. I suspect that this could be in part due to participants underestimating the difficulty of some of the search tasks. The experiment as a whole could be deceptive in difficultly to some participants. All participants begin with a simple practice search task to find a scroll saw on a grid presentation. Tasks such as these can be quite simplistic, but tasks like finding the Husky 16 oz. hammer require the participants to look at more than just the provided image. Failure to read the name of the item could result in choosing the Husky 20 oz. hammer by mistake.

One interesting observation made in the study was that the search tasks that involved items with colors that stand out had noticeably shorter fixation durations than the items that have similar appearances. This most likely means that participants treated the search tasks, like the red Rigid X4, as color matching tasks. Many of the participants avoided reading the name of the item entirely and simply matched the picture that they saw on the grey screen. On the other hand, the participants usually read the name of the item and had longer fixations when they were tasked with finding items of similar appearances. This opens the interesting question of what the results of the study would have looked like if all of the target items were similar in color and appearance to the point where participants would be much more inclined to read the names of the items. For instance, the stimuli could have been a search for "DeWalt drill", rather than just "drill." If all of the target items appeared to be visually similar, participants would have to more thoroughly search through the results to find the target item.

As for possible concerns with this study, there is the fact that the item presentation stimuli each appear multiple times for each participant. For instance, a participant can be asked to find an item on one image and then that same image can appear later with a different search task. Theoretically, a participant with a photographic memory could skew the results by remembering the locations of items in a stimuli. With that in mind, the results do not seem to necessarily suggest that that outcome occurred with any participant.

Conclusion

This study's goal was to investigate how online shoppers find the item they are looking for in a page of search results. Despite the disappointing inconclusiveness of the results in regards to the list and grid presentations, we managed to lay a foundation for future studies to build on. We now have evidence that target items are found much more quickly if they have an appearance that stands out from the other search results. Future studies will have to determine how they want to control for this factor or if it is truly necessary to control for it at all.

There is still plenty of work to be done in the field of eye tracking if we are to better understand how shoppers view search results. This study poses questions about that could be answered in future studies. One of the most intriguing of those questions is in regards to how an online shopper would view a list or a grid of very similar items, such as a list comprised entirely of yellow DeWalt drills. A study of that nature could most likely produce a more accurate analysis of the grid and drill presentations. It would also potentially be more realistic, as we suspect that many online shoppers would be more likely to type in a more descriptive search into a search engine than just "drill" or "hammer."

The ultimate goal of this study, and studies like it, is to not only improve the effectiveness of retailers, but to improve the consumer's experience as well. As technology and our understanding of consumer patterns continues to advance, we will have better and better ideas about how to present items to consumers online and in brick and mortar outlets as well.

References

Chandon, Pierre, J. Wesley Hutchinson, and Scott H. Young. *Unseen is unsold: Assessing visual equity with commercial eye-tracking data*. Insead, 2002.

Cutrell, Edward, and Zhiwei Guan. "What are you looking for?: an eye-tracking study of information usage in web search." *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 2007.

Guan, Zhiwei, and Edward Cutrell. "An eye tracking study of the effect of target rank on web search." *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 2007.

Lee, Uichin, Zhenyu Liu, and Junghoo Cho. "Automatic identification of user goals in web search." *Proceedings of the 14th international conference on World Wide Web*. ACM, 2005.

Wang, Qiuzhen, et al. "An eye-tracking study of website complexity from cognitive load perspective." *Decision support systems* 62 (2014): 1-10.