

# Users’ Perception of Search-Engine Biases and Satisfaction

Bin Han, Chirag Shah, and Daniel Saelid

University of Washington, Seattle, Washington, USA

**Abstract.** Search engines could consistently favor certain values over the others, which is considered as biased due to the built-in infrastructures. Many studies have dedicated to detect, control, and mitigate the impacts of those biases from the perspectives of search engines themselves. In our study, we pitched the perspective from end-users to analyze their perceptions of search engine biases and their satisfaction when the biases are regulated. In the study, we paired a real page from search engine Bing and a synthesized page with more diversities in the results (i.e. less biased). Both pages show the top-10 search items given search queries and we asked participants which one they prefer and why do they prefer the one selected. Statistical analyses revealed that overall, participants prefer the original Bing pages and the location where the diversities are introduced is significantly associated with users’ preferences as well. We found out that users prefer results that are more consistent and relevant to the search queries. Introducing diversities undermines the relevance of the search results and impairs users’ satisfaction to some degree. Additionally, users tend to pay more attention to the top portion of the results rather than the bottom ones.

**Keywords:** fairness · search engine bias · survey study

## 1 Introduction

Search engines often present results that are biased toward one subtopic, view, or perspective due to the way they compute relevance and measure user satisfaction. Among various types of search engine biases, one describes the case where the search engines embed features that favor certain values over the others [1, 3]. Many studies have attempted to detect, measure and mitigate the impacts generated from search engines biases, with the goal of improving users’ satisfactions. All those works aimed to address the issues from the source of the biases — search engines themselves.

In the previous study we conducted [under review], we took a different path to inspect the problem from the aspect of end-users. We paired a real search page and a synthesized page (more varieties in the search results, thus less biased) and asked participants which one they prefer. The results showed no significant difference in the ratios of selecting two pages. However, what remained unknown to us is that why did participants picked up the ones they prefer? What is the

reasoning underneath their preferences? Therefore, we revisited this study and improved our survey design catering to our goals (more details in Section 3). We would like to evaluate users’ perceptions of the biases, thus hoping to reveal the reasoning of their selections of pages. Additionally, we are interested in studying the effects on users’ satisfactions when the biases are abated.

## 2 Background

As mentioned in the introduction, copious studies have attempted to disclose and regulate search engine biases from the original sources – engines themselves. Souneil [4] discussed the political polarization caused by search engine biases and they proposed a new technique to mitigate the impacts from the biases by providing several viewpoints of the search results. Robert et al. [5] ran three experiments to study the search engine manipulation effect and deployed a browser extension to set off bias alerts from search engine manipulation effect. In addition, Kulshrestha et al. [6] proposed a framework to quantify the political bias in the search results and segment the sources where the bias comes from.

However, fewer studies have devoted to probe users’ consciousness towards the biases and their actions/behaviors associated with their awareness. Additionally, many studies (three studies exemplified above) focuses on certain domains, such as politics, where the search engine biases could have negative impacts on external events, such as election. Fewer studies have analyzed how users’ satisfactions are related with the biases in general. Consequently, our work aims to dive deeper in this direction.

## 3 Method

In this section, we present the details of the algorithm to generate the synthesized search pages. We also enunciate the two core questions we want to address from this study and three hypotheses that we would like to test on. In addition, we discuss the specifics of the survey design. Notice that in sections below, “diversities”, “varieties” and “differences” are equivalent. Introducing diversities/varieties/differences could potentially weaken the biases. “Documents”, “items” and “results” are equivalent, as they all mean the search engine results.

### Algorithm to Generate Synthesized Pages

To generate the synthesized search pages considering fairness, we implemented **statistical parity** with **epsilon-0.3** algorithm. In simple words, given a search query, we replace three items from top-10 ranked documents with three lower ranked documents (s.t. more diversities/varieties), proportionally to the frequencies of different clusters in top-10. For instance, if top-10 has eight items from cluster one and two from cluster two, we replace three out of eight items from cluster one at top-10 with three lower ranked documents from cluster two.

The algorithm is presented in Table 1 in the Appendix. It replaces three documents from top-5 in top-10 with lower ranked results. The same algorithm applies when we replace three results from bottom-5 in top-10. Please refer to Figure 1 in the Appendix for details.

### Survey Questions & Hypotheses

The two core study questions we would like to answer are:

- *Why do participants choose the one they prefer?*
- *Does introducing more varieties in the search engine results, equivalently less biased, hinder users' satisfaction?*

We raised three actionable hypotheses to potentially answer the questions:

- **H1: People do not care/notice the minute differences between the two search results:** even though we introduced lower ranked results into the top list to add varieties, the differences might not be drastic enough for some participants to notice. Or the participants might realize the differences but they do not care about which one is better.
- **H2: The location where the differences present matters. When differences are at the bottom of the search list, people do not care:** intuitively, users might treat the top-ranked results more seriously than the lower-ranked ones. Even in top-10 ranked documents, the top-5 might attract different attention than the bottom-5. Therefore, in our survey design, we put the differences in both locations (top-5 or bottom-5 in top-10 results).
- **H3: People prefer results with high relevance as opposed to high diversity:** this hypothesis could answer the second question. Introducing lower ranked search items means adding more diversities into the results, thus weakening the potential biases of search engines that consistently favor some values over the others. Unavoidably, however, adding lower ranked results would sabotage the relevance of the search results, leading to consequences of potentially lowering users' satisfactions. Therefore, we want to see whether they prefer higher relevance (more biased) or higher diversity (less biased).

### Experimental Design

The experiment starts with a consent form to be signed, which is followed by some demographic questions (e.g. age group, gender and education background. See Figure 3 in Appendix). Then the participants are provided with instructions on how to complete the survey through a quick demo (Figure 4 in Appendix). Once they are familiar with the details, they may proceed to answer the questions. The survey has 20 rounds in total. Each round consists a pair of real Bing search page and a synthesized page using the algorithm aforementioned, given a specific search query. Participants have 30 seconds to read the query, compare the items between two pages, and make a selection. Out of 20 rounds,

we randomly select 10 rounds to perform the top-5 in top-10 replacement, while the rest conducts bottom-5 in top-10 replacement. After each round, there is a reflection question (Figure 2 in Appendix) on why do they make such a choice:

- “*I did not notice any differences*” addresses the H1. The differences might not be palpable enough for the participants.
- “*I noticed some differences but did not care. So I randomly picked up one.*” addressed the H1. Participants might detect the discrepancies, but they do not make a difference in users’ satisfaction.
- “*I noticed some differences and picked the one that had more results on the same topic.*” & “*I noticed some differences and picked the one that had more variety in the results.*” They together address H3. More results on the same topic means that the documents are more consistent with each other. More variety in the results represents the introduced lower ranked results.

## 4 Results

With 137 participants completing the survey, we recorded 2,408 responses. After removing invalid entries, such as users that did not complete the survey or responses with empty selections, 111 participants with 2,134 responses were left. The basic demographic information is presented in Table 2 in the Appendix.

### Selection Ratios

Starting with the overall selection ratios of the two pages, 53.3% (N=1137) of the responses prefer the real search pages, while 46.7% (N=997) selected the synthesized versions. We ran Chi-Square goodness of fit test, where the null hypothesis states that the expected frequencies of selecting both choices are the same. The results turned out to be significant at 0.01 significance level ( $\mathbf{p=0.002}$ ). In the bottom-5 replacement group (N=1066), half responses chose the real pages and half chose the synthesized ones. There is no difference in the selection ratios. However, we observed significantly different results in the top-5 replacement group (N=1068), where 56.6% (N=604) responses preferred the real pages while 43.4% (N=464) liked the synthesized pages better. Goodness of fit test yield significant result ( $\mathbf{p<1e-4}$ )

Based on the separate tests in each replacement group, it seems that the location where the diversities present have an impact on users’ preferences. To further confirm the conjecture, we ran Chi-square test of independence on two categorical variables: users’ preferences (real page or synthesized page) and replacement group (top-5 or bottom-5). The result is significant given  $\mathbf{p=0.003}$ . It demonstrates that the location is associated with participants’ preferences.

### Reasoning Analysis

The default four reasons, corresponding to the four choices in order, are “No Diff”, “Diff Random”, “Diff Same Topic”, and “Diff Variety”. We probed the

reasons for three selection groups separately – the group that selected the real pages (called “original” group), the group that selected the top-5 replacement pages, and the group that selected the bottom-5 replacement pages. We only presented the analysis of the four default answers here because users’ own explanations are diverse and sparse, which will be analyzed in the discussion section.

The distributions of default answers for each group are exhibited in Table 3 in the Appendix. We noticed that within each group, “Diff Same Topic” dominated all other answers. Within each group, we ran Chi-square goodness of fit test, in which the null hypothesis states that the expected frequencies of the default choices are the same. All three p-values are extremely small, indicating that the observed frequencies are significantly different from the expected ones.

## 5 Discussion

From the analysis above, we showed that the proportion of participants preferring the real pages is significantly higher than that of the participants that selected the synthesized pages. Bringing up lower ranked results into the top ranked list introduces more varieties and values in the search results, thus weakening the biases of search engine in favoring certain values. However, it potentially damages the relevance and consistence of the results to the queries. Based on the result, it is reasonable to conjecture that users’ satisfactions are impaired due to the added varieties, even though the bias is mitigated.

A more interesting observation lies on the results of the two replacement groups. We hypothesized that the site where the varieties are placed play a role in affecting users’ preferences. Our results demonstrated that when the differences exhibit at bottom-5, participants had no compelling tendency of picking up one over the other. However, when they noticed the differences at top-5, they had a significant inclination to choose the real Bing page. This could be explained by [2]. It articulates that “most people use very short, imprecise search terms and rarely click beyond the first page of results. They also tend to believe that what comes at the top of the search results must be the best result. (p.13)” Kulshrestha et al. [6] also mentioned that the top-ranked results are capable of shaping users’ opinions about the topics, which demonstrates the importance of the search results’ location. In our case, we could interpret it as when the differences are shown at top-5, participants will pay more attention to pinpoint the differences and make a selection. The phenomenon was also observed in some users’ explanations in the reflection questions. Some mentioned that “*I like the first 3 results*”, “*... at the top that explained ...*”, “*preferred the top X list*”.

As for the analysis of reasoning, when participants picked up the real pages, 40% of the reasons is “Diff Same Topic”. It means that they did notice some differences between the two pages and they prefer the one with more homogeneous results. Interestingly, for users that gave their own explanations, many mentioned that the original pages provide more relevant and reliable results than the synthesized ones, which is what we expected. Introducing diversities by replacing higher ranked results with lower ranked ones will reduce biases,

but potentially hinder the relevance of the search results, thus sabotaging users' satisfaction as the aftermath.

However, when we applied the same reasoning analysis on the two replacement groups, the results do not make logical sense even they are significant. Most of the participants still selected "Diff Same Topic" as the reason, even though they picked up the synthesized pages that have more varieties rather than consistency. It means they believed that the real page are more diverse in terms of results. This could be contributed to two reasons: (1) the lower ranked results are similar to those replaced higher ranked items such that the participants did not notice the diversities; and (2) the definition of similarity and diversity on a topic is not unified and are different from each participant. Consequently, they may pick up the one that contain objectively similar results from their perspective, even though the page is subjectively more diverse. Nothing special was observed from participants' own explanations in the reflection question.

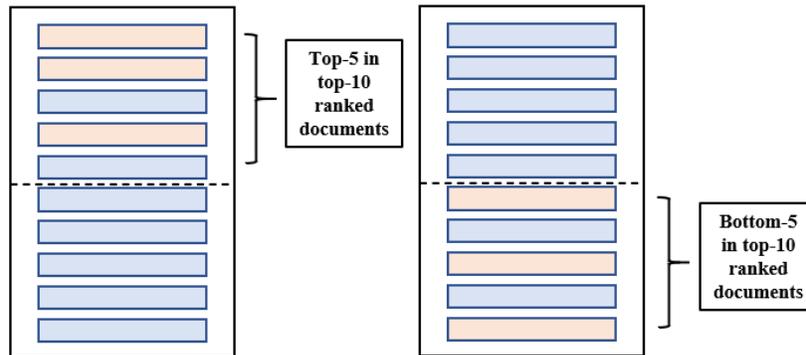
## 6 Conclusion

In our study, we designed a survey to assess users' perceptions of search engine biases, with the goal of diagnosing the reasoning underneath their preferences over the real search pages or the synthesized pages. We also investigated the effects of bias-mitigation on users' satisfactions. We noticed that overall, participants prefer the real search pages over the synthesized ones with a significant higher ratio. It shows that adding more varieties makes the results less biased but less relevant and consistent to the queries, which hurts users' satisfactions. In addition, when the diversities in the synthesized pages are present at the top-5, participants tend to prefer the real pages. However, when they are at bottom-5, there is no significant difference between the ratios of selecting two pages. It confirms our hypothesis that the location where the bias-mitigation happens is critical.

## References

1. Tanavi, H.: Search Engines and Ethics. Metaphysics Research Lab, Stanford University (2020)
2. Couvering, JV E: Search Engine Bias - The Structuration of Traffic on the World-Wide Web (2009). PhD dissertation, London School of Economics and Political Science
3. Goldman, E.: Search Engine Bias and the Demise of Search Engine Utopianism. *Yale Journal of Law and Technology*, **8**, 188–200 (2005)
4. Souneil, P. et al.: NewsCube: Delivering Multiple Aspects of News to Mitigate Media Bias. *CHI '09*, 443–452 (2009). DOI:10.1145/1518701.1518772
5. Robert, E. et al.: Suppressing the Search Engine Manipulation Effect (SEME). *Proc. ACM Hum.-Comput. Interact.*, **1**(22) (2017). DOI:10.1145/3134677
6. Kulshrestha, J. et al.: Search bias quantification: investigating political bias in social media and web search. *Information Retrieval Journal*, **22**, 188–227 (2019). DOI:10.1007/s10791-018-9341-2

## Appendix



**Fig. 1.** On the left figure, we replace three documents (marked in orange) from top-5 in top-10 ranked documents with lower ranked results. On the right figure, we replace three documents from bottom-5 in top-10 ranked documents with lower ranked results.

### Reflection

**Why do you prefer the selected page?**

- I did not notice any differences.
- I noticed some differences but did not care. So I randomly picked up one.
- I noticed some differences and picked the one that had more results on the same topic.
- I noticed some differences and picked the one that had more variety in the results.
- None of the above. Please explain briefly:

**Fig. 2.** The reflection question. After each round, participants will be asked with this mandatory reflection question. If “None of the above” is selected, the text box is mandatory as well.

## User Information Questionnaire

**What is your age?**

- 18 - 21 years old
- 22 - 28 years old
- 29 - 35 years old
- 36 - 49 years old
- 50 - 65 years old
- 65+ years old
- Prefer not to say

**What gender do you identify as?**

- Male
- Female
- Non-binary/third gender
- Others
- Prefer not to say

**What is the highest degree or level of education you have completed?**

- Some High School
- High School
- Bachelor's Degree
- Master's Degree
- Ph.D. or similar
- Trade School
- Prefer not to say

Fig. 3. Demographic questions from the survey.

### Instructions

1. You will have 30 seconds to choose between the two sets of results.
2. Select the set of results you prefer for the search topic. Selections are enabled after 3 seconds.
3. Press submit and fill out the following screen.
4. You will then be directed to the next pair of results for a next search term.
5. There will be 20 questions. In order to have a successful trial, be sure to make a choice for at least 10 questions and the associated surveys.

Here is a quick DEMO:

#### Search Result Survey

Some search terms

Search:

**Search result 1**

**WWE Stomping Grounds results: Rollins saved by Lynch, Retain...**  
<https://www.espress.com/sports/wwe/143434/WWE-Stomping-Grounds-LIVE-results-Rollins-Kingston-Lynch-Lesnar-June-23-2019> - WWE STOMPING GROUNDS results - Express Sport brought you comprehensive coverage of the show with Seth Rollins, Becky Lynch, Roman...

**WWE Stomping Grounds Full Results - Yahoo! Sports**  
<https://sports.yahoo.com/wwe-stomping-grounds-full-results-22564702.html>  
 June 23, 2019 - WWE's latest pay-per-view event, Stomping Grounds, is taking place Sunday night at the Tacoma Dome in Tacoma, Wa., and Yahoo Sports will...

**WWE Stomping Grounds PPV Live Ring and Match Results - IGN**  
<https://www.ign.com/articles/2019/06/23/wwe-stomping-grounds-live-ring-and-match-results-becky-lynch-roman-reigns>  
 June 23, 2019 - Join us for all the action at WWE Stomping Grounds as Seth Rollins, Kofi Kingston, and Becky Lynch all defend their championships!

**WWE Stomping Grounds results: live streaming match coverage...**  
<https://www.espress.com/wwe/2019/06/23/wwe-stomping-grounds-results-live-streaming-match-coverage>  
 June 23, 2019 - Get complete results and live coverage of every match on tonight's WWE Stomping Grounds PPV from Tacoma, Stomping Seth Rollins on...

**Match Star Rollins For WWE Stomping Grounds - WhatCulture.com**  
<http://whatculture.com/wwe/Match-Star-Rollins-For-WWE-Stomping-Grounds>  
 June 24, 2019 - On 'WWE Super ShowDown II', it was essentially forced throughout the build. And what a shty build it was, promoted with a TV hook as...

**Search result 2**

**WWE Stomping Grounds - Wikipedia**  
[https://en.wikipedia.org/wiki/WWE\\_Stomping\\_Grounds](https://en.wikipedia.org/wiki/WWE_Stomping_Grounds)  
 Stomping Grounds was a professional wrestling pay-per-view and WWE Network event produced by WWE for their Raw, SmackDown, and 201 Live brands.

**WWE Stomping Grounds 1 WWE**  
<https://www.wwe.com/shows/wwe-stomping-grounds>  
 WWE Stomping Grounds 2019 June 23 on WWE Network Carlini to overcome challenge Rollins at WWE Stomping Grounds WWE Stomping Grounds.

**WWE Stomping Grounds 2019 WWE**  
<https://www.wwe.com/shows/wwe-stomping-grounds/2019>  
 At the first-ever WWE Stomping Grounds pay-per-view, Becky Lynch followed spher Raw Women's Championship victory over Lacey Evans by holding.

**2019 WWE Stomping Grounds card, start time, matches, date...**  
<https://www.sportspk.com/wwe/wwe-stomping-grounds-card-start-time-matches-date-location-venue>  
 June 23, 2019 - In a few short hours, WWE will make history by presenting the Stomping Ground event for the first time in history. On the back of the...

**WWE Stomping Grounds Kickoff, June 23, 2019 - YouTube**  
<https://www.youtube.com/watch?v=22z29W9R0AQ>  
 Out of the last breaking news, analysis from a panel of experts, and kickoff matches before all...

30 seconds countdown

selection button         selection button

Fig. 4. The interface of the survey. The query shows in the search box at the top. The paired pages consist of a real Google search page and a synthesized one. Participants can select whichever one they prefer and submit.

**Algorithm**

- 
1. Group top-100 search documents into two clusters.
  2. Remove top-10 documents from the cluster lists.
  3. Calculate cluster 1 ( $c_1$ ) and cluster 2 ( $c_2$ ) frequencies ( $freq$ ) in top-10 documents.
  4. For three iterations:
    - if**  $freq_{c_1} == freq_{c_2}$  :
      - randomly select an index in the top-5 and swap with the next highest ranked document from the cluster, from which the current document was taken. Remove this document from the cluster list
    - else if**  $freq_{c_1} > freq_{c_2}$  :
      - randomly select an index in the top-5 containing a  $c_1$  document, and swap it with the next highest ranked document from  $c_2$ . If there are no remaining documents in  $c_2$ , swap with a  $c_1$  document. Remove the swapped document from cluster list
    - else if**  $freq_{c_1} < freq_{c_2}$  :
      - randomly select an index in the top-5 containing a  $c_2$  document, and swap it with the next highest ranked document from  $c_1$ . If there are no remaining documents in  $c_1$ , swap with a  $c_2$  document. Remove the swapped document from cluster list
- update frequencies  
 remove the index that has been swapped in the top-5
- 

**Table 1.** Algorithm to generate synthesized pages**Table 2.** Basic demographic information of participants — age groups, gender, and education background.

Features	Groups	Count
<b>Age Group</b>	22-28	571 (27%)
	29-35	681 (32%)
	36-49	488 (23%)
	50-65	355 (17%)
	Prefer Not To Say	39 (2%)
<b>Gender</b>	Male	1212 (57%)
	Female	863 (40%)
	Non-binary	20 (1%)
	Prefer Not To Say	39 (2%)
<b>Education</b>	High School	158 (7%)
	Bachelor	1264 (59%)
	Master	653 (31%)
	PhD (or similar)	20 (1%)
	Prefer Not To Say	39 (2%)

**Table 3.** Distributions of four default choices in each selection group. p-values are from Chi-square goodness of fit test within each group.

Groups	No Diff	Diff Random	Diff Same Topic	Diff Variety	p-value
Original	117	222	461	310	2.2e-49
Top-5	49	87	186	132	7.5e-20
Bottom-5	55	104	218	147	1.5e-23