

## Research Paper ■

# Can Cognitive Biases during Consumer Health Information Searches Be Reduced to Improve Decision Making?

ANNIE Y.S. LAU, PHD, ENRICO W. COIERA, MBBS, PHD

**Abstract Objective:** To test whether the *anchoring* and *order* cognitive biases experienced during search by consumers using information retrieval systems can be corrected to improve the accuracy of, and confidence in, answers to health-related questions.

**Design:** A prospective study was conducted on 227 undergraduate students who used an online search engine developed by the authors to find health information and then answer six randomly assigned consumer health questions. The search engine was fitted with a baseline user interface and two modified interfaces specifically designed to debias anchoring or order effect. Each subject used all three user interfaces, answering two questions with each.

**Measurements:** Frequencies of correct answers pre- and post- search and confidence in answers were collected. Time taken to search and then answer a question, the number of searches conducted and the number of links accessed in a search session were also recorded. User preferences for each interface were measured. Chi-square analyses tested for the presence of biases with each user interface. The Kolmogorov-Smirnov test checked for equality of distribution of the evidence analyzed for each user interface. The test for difference between proportions and the Wilcoxon signed ranks test were used when comparing interfaces.

**Results:** Anchoring and order effects were present amongst subjects using the baseline search interface (anchoring:  $p < 0.001$ ; order:  $p = 0.026$ ). With use of the order debiasing interface, the initial order effect was no longer present ( $p = 0.34$ ) but there was no significant improvement in decision accuracy ( $p = 0.23$ ). While the anchoring effect persisted when using the anchor debiasing interface ( $p < 0.001$ ), its use was associated with a 10.3% increase in subjects who had answered incorrectly pre-search, answering correctly post-search ( $p = 0.10$ ). Subjects using either debiasing user interface conducted fewer searches and accessed more documents compared to baseline ( $p < 0.001$ ). In addition, the majority of subjects preferred using a debiasing interface over baseline.

**Conclusion:** This study provides evidence that (i) debiasing strategies can be integrated into the user interface of a search engine; (ii) information interpretation behaviors can be to some extent debiased; and that (iii) attempts to debias information searching by consumers can influence their ability to answer health-related questions accurately, their confidence in these answers, as well as the strategies used to conduct searches and retrieve information.

■ *J Am Med Inform Assoc.* 2009;16:54–65. DOI 10.1197/jamia.M2557.

## Introduction

The use of information retrieval technologies to search for evidence that can support decision-making in healthcare has rapidly permeated many aspects of clinical practice<sup>1</sup> and become commonplace among consumers.<sup>2</sup> While much research focuses on the design of retrieval methods that can

identify potentially relevant documents, there has been little examination of the way that these retrieved documents then impact decision-making.<sup>3,4</sup>

We know that human beings seldom follow a purely rational or normative model in decision-making and are prone to a series of decision biases.<sup>5</sup> Decision-making research has for a long time identified that these biases can have adverse impact on decision outcomes.<sup>5–8</sup> Medical practitioners display cognitive biases when making clinical decisions<sup>6,7</sup> and interpreting research evidence.<sup>8</sup> Our prior work has shown that healthcare consumers and clinicians can also experience these cognitive biases whilst searching for information, and that such biases may have a negative impact on post-search decisions.<sup>9–11</sup>

However, there is little to no research to indicate whether there are mechanisms that can reduce or eliminate cognitive biases during search and modify their subsequent impact on decision-making. This paper describes two Web search interfaces designed to debias consumer health decisions,

Affiliation of the authors: Centre for Health Informatics, University of New South Wales, Sydney, Australia.

Supported by an Australian Research Council SPIRT grant and APAI scholarship C00107730, and by Australian Research Council Discovery Grant DP0452359.

The authors thank Dr David Thomas and Dr Ilse Blignault for their assistance in subject recruitment and the development of the healthcare consumer case scenarios.

Correspondence: Annie Y.S. Lau, Centre for Health Informatics, University of New South Wales, UNSW Sydney NSW 2052, Australia; e-mail: <a.lau@unsw.edu.au>.

Received for review: 07/13/07; accepted for publication: 09/30/08.

and tests whether using a search engine with these interfaces can reduce the impact of the anchoring and order biases, and thus improve post-search decision quality.

## Background

Biased decisions occur when an individual's cognition is affected by "contextual factors, information structures, previously held attitudes, preferences, and moods."<sup>12</sup> Cognitive biases arise because of limitations in human cognitive ability to properly attend to and process all the information that is available.<sup>13</sup>

In this study, we focus on anchoring and order biases. The anchoring effect, first discussed by Tversky and Kahneman,<sup>14</sup> occurs when a prior belief has a stronger than expected influence on the way new information is processed and new beliefs are formed<sup>15</sup> and has been shown to affect clinician and patient judgments.<sup>16–18</sup> Studies have reported mixed success rate in debiasing the anchoring effect. A study conducted by Lopes found that anchoring effect can be reduced by training subjects to anchor on the most informative sources rather than the initial stimuli that may not be informative.<sup>19</sup> However, a decision support system trialed by 131 university students did not mitigate the anchoring effect in a house value appraisal task.<sup>20</sup> In fact, studies have shown that undesirable behaviors may be exacerbated by attempts intended to correct them.<sup>21–25</sup> For example, Sanna and colleagues tested the strategy of thinking about alternative outcomes to minimize hindsight bias but found that the strength of the hindsight bias actually increased rather than diminished.<sup>25</sup>

An order effect occurs when the temporal order in which information is presented affects a final judgment<sup>26</sup> and can be subdivided into the primacy and recency effects. With primacy, an individual's impressions are more influenced by earlier information in a sequence.<sup>27,28</sup> With recency, impressions are more influenced by later information.<sup>29,30</sup> Although the definitions of anchoring and primacy effect have been used interchangeably in some studies,<sup>31</sup> with primacy sometimes thought of as anchoring on any belief from evidence previously obtained, primacy effect in this study strictly refers to the order in which information was presented.

Studies have shown that medical practitioners arrive at different diagnoses when the same information is presented in a different order;<sup>32–35</sup> and that the order of information presen-

tation can influence a patient's interpretation of treatment options.<sup>36</sup> Some studies have also reported success in debiasing the order effects of primacy and recency. Ashton and Kennedy conducted a study with 135 accounting auditors and found that the use of a self-review can successfully debias recency effect in evaluating the performance of a company.<sup>37</sup> Lim and colleagues showed that multimedia presentations can reduce the influence of first impression bias in a task that involved 80 university subjects evaluating the performance of an authority figure.<sup>38</sup>

## Debiasing Interfaces

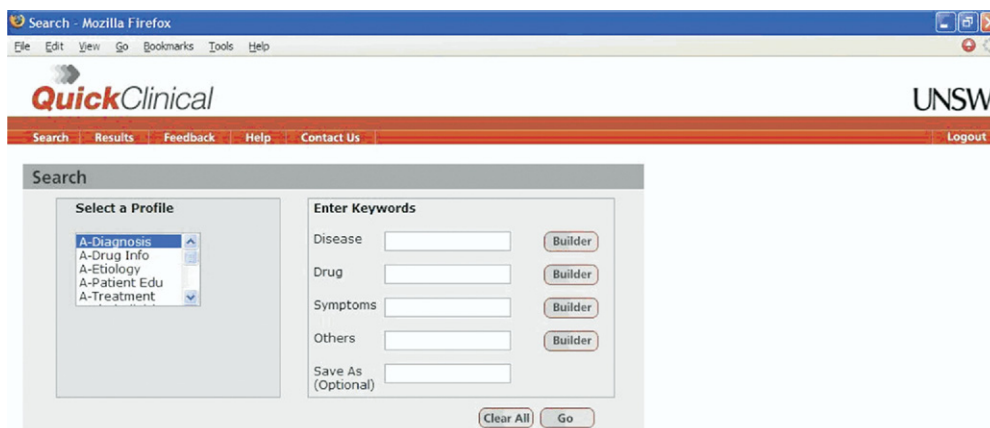
*Debiasing* has been described as any intervention that assists people to eliminate or reduce the impact of cognitive biases on their decisions and to focus their awareness on understanding the sources of their cognitive limitations.<sup>15</sup> In decision support research, many strategies and user interfaces have been proposed to debias different decision-making tasks.<sup>20,37–50</sup> Examples in the healthcare domain include assisting students to learn disease diagnosis<sup>45</sup> and providing suggestions for clinicians to avoid biases in making clinical decisions.<sup>46,47</sup> However, most attempts to debias decision-making have either not been formally evaluated, or have not produced statistically significant results when attempting to reduce the impact of biases or improve decision outcomes.<sup>7,20,21,39,40,46,47,50</sup> While many search user interfaces are designed to assist people select, process and integrate information,<sup>51</sup> we have been unable to locate studies that attempt to integrate debiasing strategies into the user interface of a search system.

### Baseline User Interface

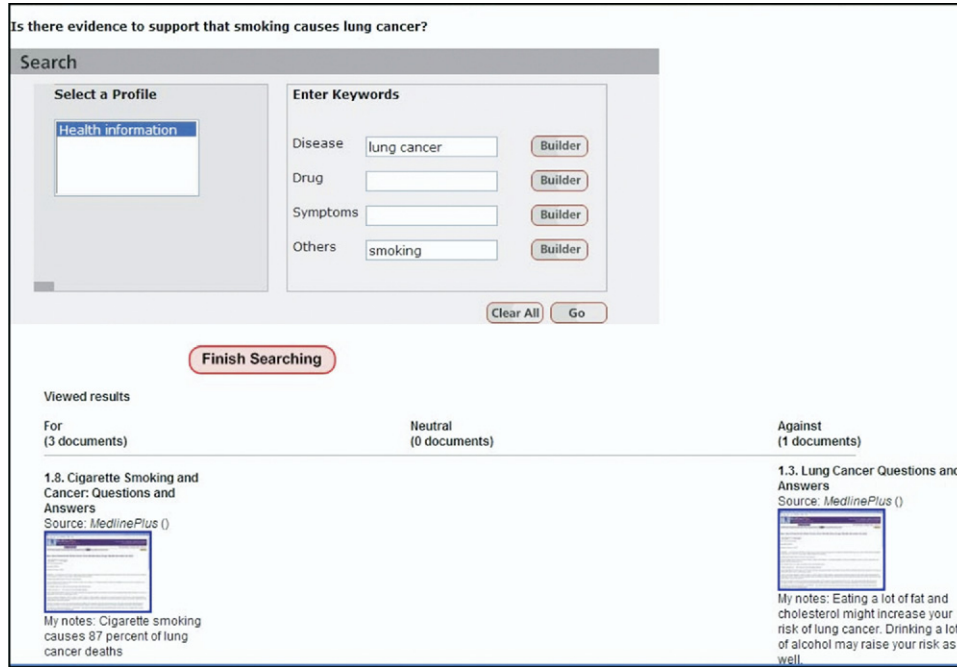
The search engine used in this study, Quick Clinical (QC), was developed at the Centre for Health Informatics, University of New South Wales.<sup>52</sup> The search engine was fitted with a baseline user interface (Figure 1), and two interfaces designed for debiasing (Figure 2, Figure 3). We have previously reported our analysis of data from the baseline user interface, which identified the presence of cognitive biases.<sup>9</sup>

### Anchor Debiasing Intervention

The anchoring effect is thought to occur when there are differences between external stimuli and subjects' interpretation of these stimuli.<sup>53</sup> Modifying the way information is presented has previously been shown to significantly improve the accuracy of decision-making,<sup>54</sup> the way people develop



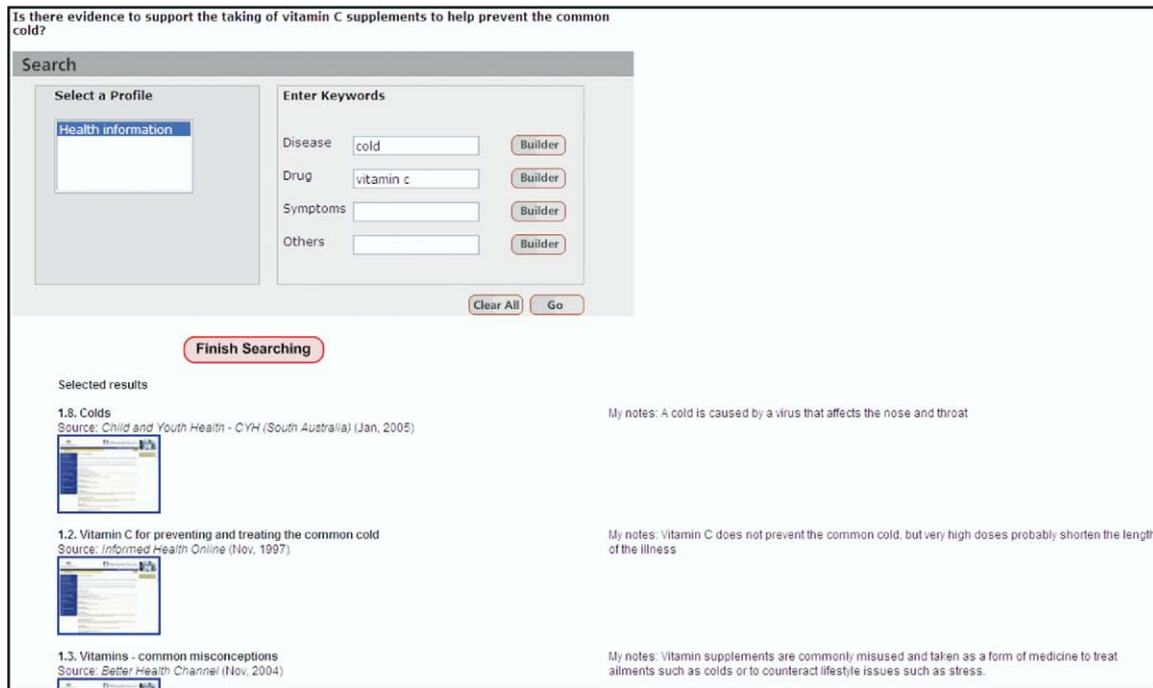
**Figure 1.** Baseline search interface (© University of New South Wales, 2004–8, used with permission).



**Figure 2.** Anchor debiasing interface—during a search session, users classify documents as for, against or neutral to the current question. Thumbnails of each document are then displayed along a *decision axis* at the bottom of the screen to show how the evidence is building up for each possible category. User notes can be associated with each document thumbnail (© University of New South Wales, 2004–8, used with permission).

concepts and make decisions,<sup>45,49</sup> as well as reducing the impact of cognitive biases due to irrelevant cues or inadequacies in judgment, such as base-rate fallacy, first impression bias, familiarity bias and confirmation bias.<sup>38,42,43,49</sup>

Our anchor debiasing interface aims to reduce the anchoring effect by asking users to assemble the body of evidence for and against a proposition, before committing to a final answer. The design rationale is that the process of assem-



**Figure 3.** Order debiasing interface—users collect and if they wish annotate documents found during the search session. These documents are displayed as thumbnails at the bottom of the screen. Upon finishing search, the thumbnails are automatically rearranged in a new sequence that attempts to minimize the impact of order effect, and users are requested to reconsider the evidence before committing to a decision (© University of New South Wales, 2004–8, used with permission).

bling evidence to answer a question will override any prior assumptions an individual may have had about the answer. Experimental psychological evidence shows that active involvement with information results in better estimates of likelihoods of events than passive engagement.<sup>55,56</sup> Further, in multi-attribute decisions, it appears that individuals will try to abstract cues from examples to form a general view of the task at hand. However, if this fails, they resort to retrieving examples from exemplar memory.<sup>57</sup> By supporting the task of cue abstraction, assisting individuals to aggregate evidence into distinct classes, they should be more likely to form a view from the current cues rather than revert to exemplars from memory.

Figure 2 displays the anchor debiasing interface which asks users to determine, after reading a document, whether the document provides evidence for, against or is neutral to a proposition. A *decision axis* at the bottom of the screen gradually builds up a collection of documents (displayed as thumbnails) that are for, neutral to, or against a proposition, as users categorize the documents they read. Users are able to capture brief notes about each document, and these are displayed below the document's thumbnail along the decision axis.

### Order Debiasing Intervention

The order effect may cause the first and/or last pieces of evidence in a sequence to be more available to recall from an individual's working memory, and thus have a disproportionate impact on a decision outcome.<sup>14</sup> The technique of self review, which asks individuals to re-examine all pieces of evidence before making a final decision, has been shown to debias the recency effect.<sup>37</sup> It has also been suggested that order debiasing methods are best applied at the evidence integration stage of the belief-updating process rather than at the evaluation of individual pieces of evidence.<sup>58</sup>

After a set of documents is retrieved and viewed, the impact of documents is shaped by an initial order bias. The proposed order debiasing interface reorders documents into a new order specifically designed to 'invert' or neutralize the first bias, by creating a counteracting order bias. The hypothesis is that the effect of combining the initial and counteracting orders will negate the two order effects. The success of this intervention depends on two things:

1. How well the counteracting order is 'shaped'—if the reordering algorithm does not accurately model the initial order bias, any attempt to transform the order may end up magnifying aspects of the initial order or introduce additional biases.
2. The relative impact of the initial and counteracting biases on decision-making—if one or the other is predominant then that may either reduce the impact of the intervention or magnify the intervention's effect and introduce a new bias.

Figure 3 displays the user interface designed to debias order effect, which again contains a tool for identifying and making notes about the documents read during search session, and collecting them as thumbnails at the bottom of the screen. At the end of the search process, the thumbnails are automatically reordered, and then redisplayed in their new sequence. Users are then invited to review these newly

sorted documents before making a post-search decision. Documents that are read in the middle of the sequence are redisplayed at first and last positions of the new sequence, and documents that are read at first and last positions are redisplayed in the middle of the new sequence. When only one document has been accessed, there is no reordering. With two documents, the order of the first and the last documents will be swapped. For searches retrieving three documents, D1, D2, and D3, the documents are rearranged randomly in either of the two arrangements: i) D2, D3, D1, or ii) D3, D1, D2. With searches retrieving larger numbers of documents, those in approximately the first, middle, and last third of the initial arrangement will be randomly rearranged, like those searches retrieving just three documents.

The rationale behind this is that if people are influenced by primacy and/or recency effects, then documents that are in the middle of the sequence would subsequently receive more attention if they get represented at the first and last positions of the new sequence.

### Hypotheses

Our study aims to test whether the two debiasing interventions described here have the potential to improve decision-making after searching for evidence. Our first hypothesis was that anchoring and order effects would be present when using the baseline search interface, but would not appear, or would be less pronounced, when users searched with the interfaces designed to debias these two effects.

Our second hypothesis was that debiasing interfaces would alter user behaviors and improve the quality of their post-search answers. We predicted that there would be differences between using the baseline and each debiasing user interface in (i) the number of questions subjects answered correctly, (ii) their confidence in these answers, (iii) the time taken to search for evidence and then answer a question, the number of searches conducted, and the number of links accessed in a search session. In addition, we hypothesized that (iv) subjects would prefer using a debiasing user interface over a baseline search interface.

### Methods

#### Prospective Experiment

Some 227 individuals were recruited from the undergraduate student population at The University of New South Wales (UNSW) and were asked to answer a set of six questions designed for healthcare consumers, with the assistance of the QC search engine provided to them. Subjects were asked to provide answers to each question before and after each search. Before commencing the study, subjects completed a ten-page online tutorial that explained the features and functionality of each user interface. They also completed an online pre-study questionnaire that collected their demographics, and a post-study questionnaire that elicited their preferences for each of the user interface.

In this study, subjects all used the baseline search interface for their first two questions. For each of the next four questions, subjects were randomly assigned to use the anchor and order debiasing interfaces, with each debiasing interface being randomly allocated twice. All six questions were randomly allocated. Each question and the expected correct answer are shown in Table 1.



Table 1 ■ Case Scenarios Presented to Subjects\*

Scenario and Question (Scenario Name)	Expected Correct Answer
1. We hear of people going on low carbohydrate and high protein diets, such as the Atkins diet, to lose weight. <i>Is there evidence to support that low carbohydrate, high protein diets result in greater long-term weight loss than conventional low energy, low fat diets?</i> (Diet)	No
2. You can catch infectious diseases such as the flu from inhaling the air into which others have sneezed or coughed, sharing a straw or eating off someone else's fork. The reason is because certain germs reside in saliva, as well as in other bodily fluids. Hepatitis B is an infectious disease. <i>Can you catch Hepatitis B from kissing on the cheek?</i> (Hepatitis B)	No
3. After having a few alcoholic drinks, we depend on our liver to reduce the Blood Alcohol Concentration (BAC). Drinking coffee, eating, vomiting, sleeping or having a shower will not help reduce your BAC. <i>Are there different recommendations regarding safe alcohol consumption for males and females?</i> (Alcohol)	Yes
4. Sudden infant death syndrome (SIDS), also known as "cot death", is the unexpected death of a baby where there is no apparent cause of death. Studies have shown that sleeping on the stomach increases a baby's risk of SIDS. <i>Is there an increased risk of a baby dying from SIDS if the mother smokes during pregnancy?</i> (SIDS)	Yes
5. Breast cancer is one of the most common types of cancer found in women. <i>Is there an increased chance of developing breast cancer for women who have a family history of breast cancer?</i> (Breast cancer)	Yes
6. Men are encouraged by our culture to be tough. Unfortunately, many men tend to think that asking for help is a sign of weakness. <i>In Australia, do more men die by committing suicide than women?</i> (Suicide)	Yes
7. Many people use home therapies when they are sick or to keep healthy. Examples of home therapies include drinking chicken soup when sick, drinking milk before bed for a better night's sleep and taking vitamin C to prevent the common cold. <i>Is there evidence to support the taking of vitamin C supplements to help prevent the common cold?</i> (Cold)	No
8. We know that we can catch AIDS from bodily fluids, such as from needle sharing, having unprotected sex and breast-feeding. We also know that some diseases can be transmitted by mosquito bites. <i>Is it likely that we can get AIDS from a mosquito bite?</i> (AIDS)	No

\*A random selection of 6 cases was presented to each subject in the study.

### Quantitative Analyses

Chi-square analysis was conducted to detect the presence of anchoring effect, compare users' confidence in their pre-search answers to each question, and detect the presence of an order effect for each user interface. The presence of an anchoring effect can be tested for by looking for a statistically significant relationship between subjects' pre-search answers and their post-search answers, and a statistically

significant relationship between subjects' confidence in their pre-search answer and their tendency to retain a pre-search answer after search (i.e., confidence in anchoring effect).<sup>9</sup>

For the order effect, the null hypothesis is that the degree of agreement or concurrence between subjects' post-search answer and the answer suggested by a document is not influenced by the position in a search session at which the document was accessed.<sup>9</sup> The alternative hypothesis is that there is greater concurrence between subjects' post-search answer and the answer suggested by documents accessed at either the first or last position, because of their disproportionate influence mediated by the primacy and recency effects.

The Kolmogorov-Smirnov (K-S) test<sup>59</sup> was used to test for equality of distribution of accessed evidence in the comparison populations, to ensure that unequal access to evidence did not bias the outcome; that is, that by chance that subjects using one of the search interfaces did not access more influential documents in answering a question than those using one of the other interfaces. To do this, we tested for equality of distribution of evidence impact, where the impact that an individual document has on a decision is modeled as an association between the frequency with which the document is accessed and the frequency with which accessing it is associated with a correct or incorrect decision. Document impact is calculated in the form of a likelihood ratio (LR).<sup>9</sup> Documents with a  $LR > 1$  are more likely to be associated with a correct answer, and those with a  $LR < 1$  are more likely to be associated with an incorrect answer to a question.

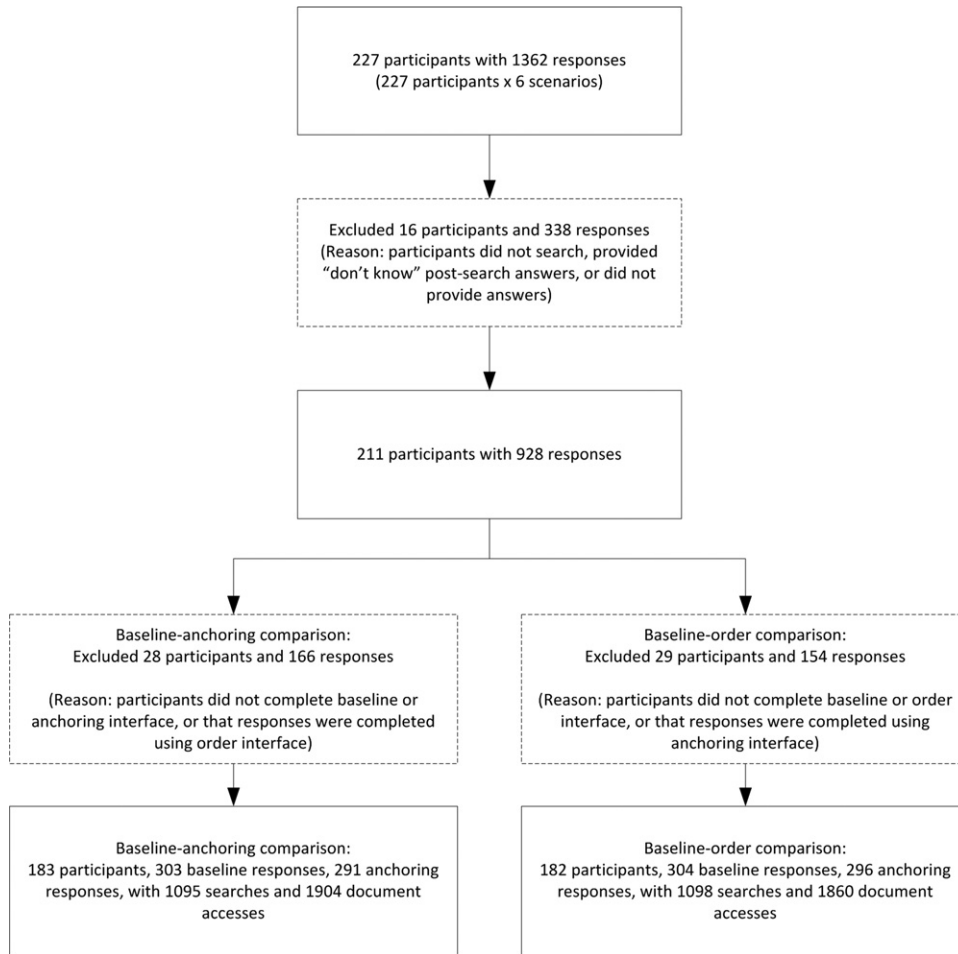
Four aspects of decision-making namely, response accuracy, user confidence, search behavior and user preference were compared between baseline search interface and each debiasing interface. The test for difference between proportions was used to compare response accuracy and users' confidence in their answers in the baseline interface with these measures in each debiasing interface. The Wilcoxon signed ranks test was used to compare differences in the amount of time taken to search and answer a question, the number of searches conducted and the number of documents accessed between different interfaces. Descriptive statistics were used to report which interface subjects found most useful, enjoyed using the most and preferred to use for future information searching.

### Results

To ensure that only valid pairwise comparisons were included in this analysis, data were excluded if a subject did not search or did not use the baseline search interface or a debiasing interface (Figure 4). After data exclusion, 183 subjects were available for the baseline-anchoring comparison. They produced 303 baseline responses, 291 anchoring responses, 1095 searches, and 1904 document accesses. Also, 182 subjects were available for the baseline-order comparison. These subjects produced 304 baseline responses, 296 order responses, 1098 searches, and 1860 document accesses.

#### Anchoring Effect and Confidence in Anchor Effect

An anchoring effect remained detectable in the responses completed with the assistance of the baseline search interface, the anchor debiasing interface and the order debiasing interface. Overall, there was a statistically significant rela-



**Figure 4.** Data exclusion for comparing the effectiveness of baseline and each debiasing user interface.

tionship between subjects’ pre-search answers and their post-search answers (indicating an anchoring effect) for all three interfaces (baseline:  $\chi^2 = 50.25$ ,  $df = 1$ ,  $p < 0.001$ ; anchor:  $\chi^2 = 22.48$ ,  $df = 1$ ,  $p < 0.001$ ; order:  $\chi^2 = 22.77$ ,  $df = 1$ ,  $p < 0.001$ ). (Table 2; also illustrated in Figure 5). Subjects’ confidence in answers were not found to have a statistically significant association with their pre-search answer retention rate using the baseline search interface ( $\chi^2 = 2.67$ ,  $df = 3$ ,  $p = 0.45$ ), but this association was found to be significant when using the anchor debiasing interface ( $\chi^2 = 9.91$ ,  $df =$

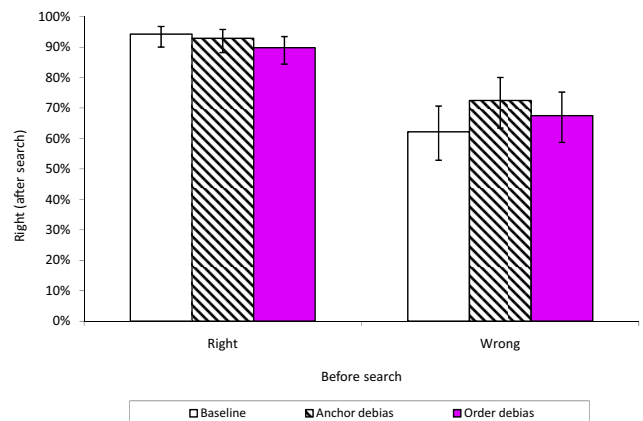
3,  $p = 0.019$ ) and the order debiasing interface ( $\chi^2 = 11.39$ ,  $df = 3$ ,  $p = 0.054$ ) (Table 3; also illustrated in Figure 6).

The K-S test showed that there were no differences in the distribution of evidence between subjects with pre-search “right” answers and subjects with pre-search “wrong” answers when they used the baseline and the anchor debiasing interfaces (baseline: K-S  $Z = 1.209$ ,  $|D| = 0.103$ ,  $p = 0.108$ ; anchor: K-S  $Z = 0.957$ ,  $|D| = 0.073$ ,  $p = 0.319$ ). However,

**Table 2** ■ Relationship between Pre-search Answer and Post-search Answer

Before Search	After Search	
	Right	Wrong
<b>Baseline search interface*</b>		
Right (n = 192)	181 (94%)	11 (6%)
Wrong (n = 111)	69 (62%)	42 (38%)
<b>Anchor debiasing interface*</b>		
Right (n = 182)	169 (93%)	13 (7%)
Wrong (n = 109)	79 (73%)	30 (27%)
<b>Order debiasing interface*</b>		
Right (n = 176)	158 (90%)	18 (10%)
Wrong (n = 120)	81 (68%)	39 (32%)

\*Anchoring effect was found to be significant.



**Figure 5.** Relationship between pre-search answer and post-search correctness.

**Table 3** ■ Relationship between Confidence in Pre-search Answer and Retention of Pre-search Answer after Searching

Confidence (Before Search)	Retained Pre-Search Answer After Searching?	
	Yes	No
<b>Baseline search interface*</b>		
Not confident (n = 33)	22 (67%)	11 (33%)
Somewhat confident (n = 71)	51 (72%)	20 (28%)
Confident (n = 85)	62 (73%)	23 (27%)
Very confident (n = 106)	84 (79%)	22 (21%)
<b>Anchor debiasing interface†§</b>		
Not confident (n = 43)	22 (51%)	21 (49%)
Somewhat confident (n = 78)	50 (64%)	28 (36%)
Confident (n = 90)	68 (76%)	22 (24%)
Very confident (n = 74)	55 (74%)	19 (26%)
<b>Order debiasing interface‡§</b>		
Not confident (n = 46)	24 (52%)	22 (48%)
Somewhat confident (n = 77)	59 (77%)	18 (23%)
Confident (n = 86)	51 (59%)	35 (41%)
Very confident (n = 85)	62 (73%)	23 (27%)

\*8 responses were excluded because they did not provide a pre-search confidence.

†6 responses were excluded because they did not provide a pre-search confidence.

‡2 responses were excluded because they did not provide a pre-search confidence.

§Confidence in anchoring effect was found to be significant.

a significant difference in evidence distribution was detected for those answers completed using the order debiasing interface ( $K-S Z = 1.426, |D| = 0.105, p = 0.034$ ). Yet, this difference was not evident in further analysis which found that subjects with pre-search right answers and subjects with pre-search wrong answers accessed statistically comparable proportions of positive evidence (i.e., evidence with  $LR > 1$ ) and negative evidence (i.e., evidence with  $LR < 1$ ) ( $\chi^2 = 1.008, df = 1, p = 0.315$ ).

**Order Effect**

Order effect remained detectable amongst responses completed using the baseline and anchor debiasing interfaces, but not amongst those using the order debiasing interface. The relationship between the access position of a document and concurrence between the post-search answer and the answer suggested by the document (i.e., order effect) was

**Table 4** ■ Relationship between Document Access Position and Concurrence between Post-search Answer and Document-suggested Answer

Access Position	Concurrence Between Post-Search Answer and Document?	
	Yes	No
<b>Baseline search interface*§</b>		
First (n = 185)	171 (92%)	14 (8%)
Middle (n = 342)	306 (90%)	36 (10%)
Last (n = 185)	155 (84%)	30 (16%)
<b>Anchor debiasing interface†§</b>		
First (n = 215)	178 (83%)	37 (17%)
Middle (n = 204)	186 (91%)	18 (8.8%)
Last (n = 215)	182 (85%)	33 (15%)
<b>Order debiasing interface‡§</b>		
First (n = 212)	173 (82%)	39 (18%)
Middle (n = 573)	481 (84%)	92 (16%)
Last (n = 212)	184 (87%)	28 (13%)

\*122 accessed documents excluded because subjects accessed only one document or a likelihood ratio could not be calculated for the document.

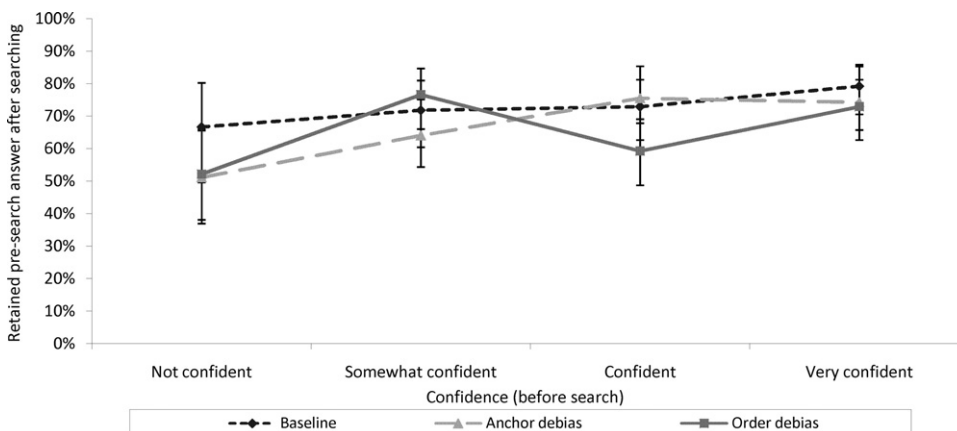
†436 accessed documents excluded because subjects accessed only one document or a likelihood ratio could not be calculated for the document.

‡29 accessed documents excluded because subjects accessed only one document or a likelihood ratio could not be calculated for the document.

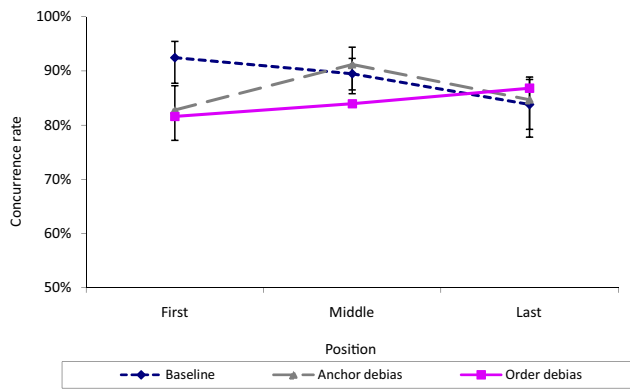
§Order effect was found to be significant.

statistically significant amongst those using the baseline search interface ( $\chi^2 = 7.27, df = 2, p = 0.026$ ) and the anchor debiasing interface ( $\chi^2 = 6.75, df = 2, p = 0.0342$ ), but not significant amongst those using the order debiasing interface ( $\chi^2 = 2.14, df = 2, p = 0.34$ ). (Table 4; also illustrated in Figure 7).

For both baseline and the order debiasing interfaces, K-S test analyses showed that there were no statistically significant differences in the distribution of evidence between first and middle positions (baseline:  $K-S Z = 0.788, |D| = 0.067, p = 0.564$ ; order:  $K-S Z = 0.527, |D| = 0.042, p = 0.944$ ), first and last positions (baseline:  $K-S Z = 0.728, |D| = 0.076, p = 0.665$ ; order:  $K-S Z = 0.826, |D| = 0.080, p = 0.503$ ), nor between middle and last positions (baseline:  $K-S Z = 1.014, |D| = 0.093, p = 0.255$ ; order:  $K-S Z = 1.008,$



**Figure 6.** Relationship between confidence in pre-search answer and pre-search answer retention rate.



**Figure 7.** Relationship between document access position and concurrence rate between post-search answer and document-suggested answer.

$|D| = 0.081, p = 0.262$ ). For the anchor debiasing interface, K-S test analyses showed that there were statistically significant differences in the distribution of evidence between first and middle positions (K-S  $Z = 1.612, |D| = 0.158, p = 0.011$ ), between middle and last positions (K-S  $Z = 1.374, |D| = 0.134, p = 0.046$ ) but not between first and last positions (K-S  $Z = 0.338, |D| = 0.033, p = 1.000$ ). Yet, subgroup analysis for the anchor debiasing interface showed that positive evidence (i.e., evidence with  $LR > 1$ ) and negative evidence (i.e., evidence with  $LR < 1$ ) distributions across the first, middle, and last positions were statistically comparable ( $\chi^2 = 0.905, df = 2, p = 0.636$ ).

**Impact on Response Accuracy**

**Baseline vs. Anchor:** Subjects who had incorrect pre-search answers, and who used the anchor debiasing interface were more likely to answer correctly post-search than those who used the baseline search interface (baseline: 62%, anchoring: 73%,  $Z = -1.64, p = 0.10$ ). Otherwise, there were no statistically significant differences in the overall number of correct post-search answers provided using baseline or the anchor debiasing interface (baseline: 83%, anchoring: 85%,  $Z = -0.90, p = 0.37$ ), nor in the number of correct post-search answers provided by those who were correct pre-search (baseline: 94%, anchoring: 93%,  $Z = 0.56, p = 0.58$ ) (Table 5).

**Baseline vs. Order:** There were no statistically significant differences between baseline and the order debiasing interface in the number of correct post-search answers (baseline:

**Table 5 ■ Comparison of Post-search Correctness in Baseline Search Interface and Anchor Debiasing Interface**

After Search	Anchor Debiasing		Z	p
	Baseline	Interface		
All responses	(n = 303)	(n = 291)		
Right	250 (83%)	248 (85%)	-0.90	0.37
Wrong	53 (18%)	43 (15%)	0.90	0.37
Right before search	(n = 192)	(n = 182)		
Right	181 (94%)	169 (93%)	0.56	0.58
Wrong	11 (5.7%)	13 (7.1%)	-0.56	0.58
Wrong before search	(n = 111)	(n = 109)		
Right	69 (62%)	79 (73%)	-1.64	0.10
Wrong	42 (38%)	30 (28%)	1.64	0.10

**Table 6 ■ Comparison of Post-search Correctness in Baseline Search Interface and Order Debiasing Interface**

After Search	Order Debiasing		Z	p
	Baseline	Interface		
All responses	(n = 304)	(n = 296)		
Right	249 (82%)	239 (81%)	0.36	0.71
Wrong	55 (18%)	57 (19%)	-0.36	0.71
Right before search	(n = 193)	(n = 176)		
Right	180 (93%)	158 (90%)	1.20	0.23
Wrong	13 (6.7%)	18 (10%)	-1.20	0.23
Wrong before search	(n = 111)	(n = 120)		
Right	69 (62%)	81 (68%)	-0.85	0.40
Wrong	42 (38%)	39 (33%)	0.85	0.40

82%, order: 81%,  $Z = 0.36, p = 0.71$ ). There were also no significant differences in the number of post-search correct answers amongst those who were correct pre-search (baseline: 93%, order: 90%,  $Z = 1.20, p = 0.23$ ), nor amongst those who were incorrect pre-search (baseline: 62%, order: 68%,  $Z = -0.85, p = 0.40$ ) (Table 6).

**Impact on Confidence**

**Baseline vs. Anchor:** Subjects using the anchor debiasing interface were more likely to report being confident or very confident in their correct post-search answers than those using the baseline search interface (baseline: 94%, 95% CI 90 to 96; anchoring: 97%, 95% CI: 94 to 99;  $Z = -1.92, p = 0.055$ ) (Table 7). There were no statistically significant differences in the number of subjects who were highly confident about their incorrect post-search answers, when comparing those using the baseline and anchor debiasing interfaces (baseline: 81%, 95% CI: 69 to 89; anchoring: 79%, 95% CI: 65 to 89;  $Z = 0.25, p = 0.80$ ).

**Baseline vs. Order:** Between baseline and the order debiasing interface, there were no statistically significant differences in the number of highly confident correct post-search answers (baseline: 94%, 95% CI: 90 to 96; order: 95%, 95% CI: 92 to 97;  $Z = -0.70, p = 0.48$ ), nor in highly confident incorrect post-search answers (baseline: 82%, 95% CI: 70 to 90; order: 84%, 95% CI: 73 to 91;  $Z = -0.34, p = 0.73$ ) (Table 8).

**Impact on Search Behavior**

**Baseline vs. Anchor:** Subjects using the anchor debiasing interface (i) took longer to search and answer a question

**Table 7 ■ Comparison of Confidence in Right and Wrong Post-search Answers between Baseline and Anchor Debiasing interface**

Post-Search Confidence	Anchor Debiasing		Z	p
	Baseline	Interface		
Right after search	(n = 250)	(n = 248)		
Not confident/ Somewhat confident	16 (6.4%)	7 (2.8%)	1.92	0.055
Confident/Very confident	234 (94%)	241 (97%)	-1.92	0.055
Wrong after search	(n = 53)	(n = 43)		
Not confident/ Somewhat confident	10 (19%)	9 (21%)	-0.25	0.80
Confident/Very confident	43 (81%)	34 (79%)	0.25	0.80



**Table 8** ■ Comparison of Confidence in Right and Wrong Post-search Answers between Baseline and Order Debiasing Interface

Post-Search Confidence	Baseline	Order Debiasing Interface	Z	p
Right after search	(n = 249)	(n = 239)		
Not confident/ Somewhat confident	15 (6.0%)	11 (4.6%)	0.70	0.48
Confident/Very confident	234 (94%)	228 (95%)	-0.70	0.48
Wrong after search	(n = 55)	(n = 57)		
Not confident/ Somewhat confident	10 (18%)	9 (16%)	0.34	0.73
Confident/Very confident	45 (82%)	48 (84%)	-0.34	0.73

(baseline: 322 seconds, standard deviation (SD): 270; anchoring: 372 seconds, SD: 295;  $Z = 2.129$ ,  $p = 0.033$ ), (ii) conducted fewer searches (baseline: 2.14, SD: 1.65; anchoring: 1.53, SD: 1.37;  $Z = -7.15$ ,  $p < 0.001$ ), and (iii) accessed more documents (baseline: 2.58, SD: 2.46; anchoring: 3.86, SD: 3.52;  $Z = -5.53$ ,  $p < 0.001$ ) than those using the baseline search interface (Table 9).

Baseline vs. Order: There were (i) no statistically significant differences in the amount of time taken to search and answer a question between baseline and the order debiasing interface (baseline: 325 seconds, SD: 271 vs. order: 332 seconds, SD: 291;  $Z = -0.57$ ,  $p = 0.57$ ). However, subjects using the order debiasing interface (ii) conducted fewer searches (baseline: 2.16, SD: 1.67; order: 1.49, SD: 0.97;  $Z = -6.81$ ,  $p < 0.001$ ), and (iii) accessed more documents (baseline: 2.62, SD: 2.45; order: 3.59, SD: 3.09;  $Z = -4.10$ ,  $p < 0.001$ ) than those using baseline search interface (Table 10).

### User Preference

The user interface subjects found most useful, enjoyed using the most and preferred to use in the future are (in descending order, from most favorable): anchor debiasing interface, order debiasing interface and baseline search interface (Table 11).

## Discussion

### Can User Interfaces Debias User Decisions after Search?

An order effect was not detected amongst subjects using the order debiasing interface. The anchor debiasing interface in

**Table 9** ■ Comparison of Search Sessions Using Baseline and Anchor Debiasing Interface

Search Attribute	Baseline	Anchor Debiasing Interface	Z	P
	Average (SD)	Average (SD)		
Time taken (seconds)	322 (270)	372 (295)	2.129	0.033
No. of searches	2.14 (1.65)	1.53 (1.37)	-7.15	<0.001
No. of documents accessed	2.58 (2.46)	3.86 (3.52)	3.522	<0.001

**Table 10** ■ Comparison of Search Sessions Using Baseline and Order Debiasing Interface

Search Attribute	Baseline	Order Debiasing Interface	Z	P
	Average (SD)	Average (SD)		
Time taken (seconds)	325 (271)	332 (291)	-0.57	0.57
No. of searches	2.16 (1.67)	1.49 (0.97)	-6.81	<0.001
No. of documents accessed	2.62 (2.45)	3.59 (3.09)	-4.10	<0.001

contrast did not remove the anchoring effect, but did lead more subjects who were incorrect in their pre-search answer to change their answer after searching, producing an increase in the number of times subjects were highly confident about their correct post-search answers. There is a complex association between confidence and anchoring,<sup>60</sup> and the effect of the anchor debiasing interface on subjects who are less confident suggests that, for this group, there may have been a reduction in the anchoring effect. Overall, although mixed, these results provide good evidence that at least some cognitive biases may be moderated by search engine user interface design.

### Can Debiasing User Interfaces Improve Decision-making after Search?

Although the anchoring effect was not completely removed, use of the anchor debiasing interface was associated with an increase in the number of subjects with an incorrect pre-search answer who correctly answer questions post-search. This may reflect the impact of the interface on the behavior of those who were less confident pre-search, since their lack of confidence may have made them more willing to consider

**Table 11** ■ User Interface Preferences Reported by Subjects (n = 183)

System Nominated by Respondent	Frequency (%)	95% Confidence Interval (CI)
Interface found most useful		
Anchor debiasing interface (i.e. for/against tool)	80 (44%)	37 to 51
Order debiasing interface (i.e. keep document tool)	75 (41%)	34 to 48
Baseline search interface	27 (15%)	10 to 21
No response	1 (0.5%)	Not applicable
Interface found most enjoyable		
Anchor debiasing interface (i.e. for/against tool)	83 (45%)	38 to 53
Order debiasing interface (i.e. keep document tool)	60 (33%)	26 to 40
Baseline search interface	39 (21%)	16 to 28
No response	1 (0.5%)	Not applicable
Interface preferred for future use		
Anchor debiasing interface (i.e. for/against tool)	77 (42%)	35 to 49
Order debiasing interface (i.e. keep document tool)	76 (42%)	35 to 49
Baseline search interface	29 (16%)	11 to 22
No response	1 (0.5%)	Not applicable

alternative answers after the search was completed. In contrast, although the order effect was not detected amongst responses completed using the order debiasing interface, there were no significant differences in the number of correct post-search answers, or in the distribution of confidence in these answers between baseline and the order debiasing interface.

Our previous work modeling the impact of biases on post-search decisions<sup>11</sup> suggests that anchoring has a very large impact on post-search decisions, but that the order effect has a relatively modest impact on decisions. Consequently, we hypothesize from these results that while the anchor debiasing interface did not eliminate the bias altogether, the interface did reduce this bias to some extent. This is reflected in the improved decision outcomes. In contrast, whilst the order effect was eliminated, the modest impact of this bias may have required a much larger sample size to detect any significant impact on decision outcome.

A number of other possibilities may explain why the order intervention did not affect decision outcome. The relatively few documents accessed per session (Table 10) may also have had an impact on the size of the order effect, and as a consequence, this also may have had an impact on decisions. Other studies have shown that document relevance judgments are not influenced by order effects when less than 15 documents are retrieved.<sup>61,62</sup> However as we have shown elsewhere, there appears to be a very weak relationship at best between relevance judgments and decision outcomes,<sup>63</sup> making it difficult to extrapolate from these experiments to our own experimental setting.

Secondly, we expected the shape of the order bias curve to be 'U' shaped,<sup>9</sup> with first and last documents having equal impact. Figure 7 shows that at baseline, the order bias we actually found was linear and stronger for earlier rather than later documents. Consequently the debiasing algorithm did not perfectly model the behavior of the initial order bias in our subjects, and may have therefore been less effective. This raises an interesting question about the innate variability of the behavior of these biases in different populations or settings, and the need to customize debiasing interfaces to local circumstances.

One additional interesting observation is that a new order effect was detected amongst subjects using the anchor debiasing intervention. This may be due to the statistically significant difference in distribution of evidence, as measured by document impact on decisions, accessed across the different order positions in this experiment. Another is that the anchor debiasing intervention has introduced an unanticipated order effect. Further work would be needed to explore whether these, or other explanations, are most likely, and underlines the complexity of the phenomena being studied.

### Do Debiasing User Interfaces Significantly Alter Search Behaviors?

Although subjects using the anchor debiasing interface took longer to search and answer a question, they conducted fewer searches and accessed more documents, suggesting the interface resulted in users spending more time considering the individual pieces of evidence they discovered during the search process. There were no significant differ-

ences in the amount of time taken to search and answer a question using the order debiasing interface, but users conducted fewer searches and accessed more documents, suggesting that the interface again encouraged users to consider more of the documents they had retrieved before searching again.

### Are Debiasing Interfaces Acceptable Alternatives to Conventional Search Interfaces?

Overall, the anchor debiasing interface was the most preferred user interface, and both debiasing interfaces were preferable to the more standard baseline interface. This suggests that Web search interfaces that assist with assembling and presenting evidence about general health questions may be well received enhancements to current Web search systems, at least for consumers seeking health information like the student subjects in this study.

### Limitations

There was a strict order of exposure to interfaces in this study, with every subject first using the baseline interface for two questions, before randomly accessing the two intervention interfaces over four questions. Familiarity with the baseline interface may have introduced a learning effect that enhanced performance with the later interfaces. The random allocation of debiasing interfaces ensured there was no learning effect associated with use of the debiasing interfaces themselves. Furthermore, learning effect was unlikely to have affected the findings for the following reasons:

- Allocation to the six questions was random across all three interfaces, so each interface was used to answer the same cohort of questions.
- Subjects were given a 10-page tutorial covering all three interfaces before commencing the study. They therefore had familiarity with the features of each interface before answering questions, but the purpose of each interface was not explained.
- If there were a learning effect in action, the accuracy of post-search answers achieved by the order debiasing interface should be higher than the accuracy of the baseline interface; however, this is not the case as demonstrated in Table 6.

Our experimental design does not make it possible to directly determine the stand alone impact of the debiasing interventions as we did not ask subjects to answer questions between retrieving documents and then using the debiasing intervention. Doing so would have introduced a further sequential judgment task into the experiment (being asked to answer the same question several times) which itself may change the opinion that is ultimately formed.<sup>64-66</sup> Introducing a measurement step in the middle of the debiasing experiment may thus have had a "Heisenberg uncertainty" effect, in that the act of measurement itself changed the outcome. For this reason we believe we adopted the better experimental design, where we infer the impact of the debiasing interfaces by comparing using them to not using them. Future work may attempt to more directly measure the impact of debiasing interfaces by introducing the sequential decision task, but also attempt to control for any confounding effect of answering the same question repeatedly. For example, one could introduce an experimental arm

where subjects answer the same question several times without exposure to a debiasing interface.

Overall, it is likely that the evidence for the presence of biases and the potential benefits of debiasing can be generalizable to different domains, as well as people with different levels of expertise. Given that biases are innate in human decision-making, it is plausible that the impact of cognitive biases on information searching is not restricted to health-related decision-making by consumers only. Physicians and patients are equally prone to errors in decision-making due to their prior beliefs or irrelevant cues.<sup>16</sup> Furthermore, the baseline interface used in this study is likely to be representative of the type of search systems in use clinically. The debiasing interfaces presented here may be more suited to reflective decision-making where time pressure is not a significant factor; thus, the impact of this kind of interface on time-constrained clinical decision-making will also need to be assessed. Studies of time-constrained searches would help to ensure that the additional time-cost of searching delivers comparable or greater benefit, measured in improved decision outcomes and clinical decision velocity.<sup>67</sup> However, there seems no reason in principle to not to ask experts to consider the evidence “as is” and check that their prior beliefs do not overly shape the way they view new evidence. With the rapid rate of change in biomedical knowledge, many areas of expertise are in constant evolution, and so, anchoring onto past beliefs may be another significant source of clinical error.<sup>68</sup>

## Conclusion

This study provides preliminary evidence that attempts to debias information searching can reduce the impact of some cognitive biases, improve decision outcomes, and alter information searching behavior. The undergraduate student subjects who used the debiasing interfaces in this study conducted fewer searches on average and accessed more documents on average to answer a question. With the lack of benchmark literature on bias and debiasing studies in information searching, further research will be needed to reinforce and extend these findings, both with healthcare consumers and with systems designed to retrieve information for expert healthcare professionals. There is still much to be learned about the impact of cognitive biases on information searching and decision-making. This research holds out the promise that, as we learn more, it will be possible to redesign information retrieval systems to significantly improve the impact of search engine design on decision-making.

## References ■

- Hersh WR. Evidence-based medicine and the internet. *ACP J Club*. 1996;125(1):A14–6.
- Eysenbach G, Jadad AR. Evidence-based patient choice and consumer health informatics in the internet age. *J Med Internet Res*. 2001;3(2):e19.
- Westbrook JL, Coiera EW, Gosling AS. Do online information retrieval systems help experienced clinicians answer clinical questions? *J Am Med Inform Assoc*. 2005;12(3):315–21.
- Hersh WR. Ubiquitous but unfinished: On-line information retrieval systems. *Med Decis Making*. 2005;25(2):147–8.
- Kahneman D, Slovic P, Tversky A. *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press; 1982.
- Elstein AS. Heuristics and biases: Selected errors in clinical reasoning. *Acad Med*. 1999;74(7):791–4.
- Croskerry P. Achieving quality in clinical decision making: Cognitive strategies and detection of bias. *Acad Emerg Med*. 2002;9(11):1184–204.
- Kaptchuk TJ. Effect of interpretive bias on research evidence. *BMJ*. 2003;326(7404):1453–5.
- Lau AYS, Coiera EW. Do people experience cognitive biases while searching for information? *J Am Med Inform Assoc*. 2007;14(5):599–608.
- Lau AYS, Coiera EW. How do clinicians search for and access biomedical literature to answer clinical questions? *MEDINFO 2007*; Brisbane, Australia. 2007.
- Lau AYS, Coiera EW. A Bayesian model that predicts the impact of web searching on decision making. *J Am Soc Inform Scie Technol*. 2006;57(7):873–80.
- de Mello GE. In need of a favorable conclusion: The role of goal-motivated reasoning in consumer judgments and evaluations. Summarised version of dissertation: University of Southern California; 2004.
- Kruglanski AW, Ajzen I. Bias and error in human judgment. *Eur J Soc Psychol*. 1983;13(1):1–44.
- Tversky A, Kahneman D. Judgment under uncertainty: Heuristics and biases. *Science*. 1974;185(4157):1124–31.
- Wickens CD, Hollands JG. *Decision making. Engineering psychology and human performance*. 3rd ed. Upper Saddle River, NJ: Prentice Hall; 2000. p. 293–336.
- Brewer NT, Chapman GB, Schwartz JB, Bergus GR. The influence of irrelevant anchors on the judgments and choices of doctors and patients. *Med Decis Making*. 2007;27(2):203–11.
- Berner E, Maisiak R, Heudebert G, Young K, editors. *Clinician performance and prominence of diagnoses displayed by a clinical diagnostic decision support system AMIA Annual Symposium Proceedings*; 2003.
- Friedlander M, Stockman S. Anchoring and publicity effects in clinical judgment. *J Clin Psychol*. 1983;39:637–43.
- Lopes LL. *Procedural debiasing*: Madison: Wisconsin Human Information Processing Program; 1982. Report No.: WHIPP15.
- George JF, Duffy K, Ahuja M. Countering the anchoring and adjustment bias with decision support systems. *Decis Supp Syst*. 2000;29(2):195–206.
- Weinstein ND, Klein WM. Resistance of personal risk perceptions to debiasing interventions. *Health Psychol*. 1995;14(2):132–40.
- Curley SP, Yates JF, Abrahms RA. Psychological sources of ambiguity avoidance. *Organ Behav Hum Decis Process*. 1986;38(2):230–56.
- Tetlock PE, Boettger R. Accountability amplifies the status quo effect when change creates victims. *J Behav Decis Making*. 1994;7(1):1–23.
- Ashton RH. Pressure and performance in accounting decision settings: Paradoxical effects of incentives, feedback and justification. *J Account Res*. 1990;28(Suppl):148–80.
- Sanna LJ, Schwarz N, Stocker SL. When debiasing backfires: Accessible content and accessibility experiences in debiasing hindsight. *J Exp Psychol Learn Mem Cogn*. 2002;28(3):497–502.
- Wang H, Zhang J, Johnson TR, editors. *Human belief revision and the order effect*. Twenty-second Annual Conference of the Cognitive Science Society; 2000; Hillsdale, NJ.
- Murdock BB. An analysis of the serial position curve. In: Roediger HLI, Nairne JS, Neath I, Surprenant AM, editors. *The nature of remembering: Essays in honor of Robert G Crowder*: American Psychological Association (APA); 2001.
- Lund FH. The psychology of belief: IV. The law of primacy in persuasion. *J Abnorm Soc Psychol*. 1925;20:183–91.
- Cromwell H. The relative effect on audience attitude of the first versus the second argumentative speech of a series. *Speech Monographs*. 1950;17:105–22.



30. Luchins AS. Primacy-recency in impression formation. In: Hovland CI, Mandell E, Campbell E, Brock T, Luchins AS, Cohen A, et al., editors. *The order of presentation in persuasion*. New Haven, CT: Yale University Press; 1957. p. 33–61.
31. Mumma GH, Wilson SB. Procedural debiasing of primacy/anchoring effects in clinical-like judgments. *J Clin Psychol*. 1995;51(6):841–53.
32. Bergus GR, Chapman GB, Gjerde C, Elstein AS. Clinical reasoning about new symptoms despite preexisting disease: Sources of error and order effects. *Fam Med*. 1995;27(5):314–20.
33. Bergus GR, Chapman GB, Levy BT, Ely JW, Oppliger RA. Clinical diagnosis and the order of information. *Med Decis Making*. 1998;18(4):412–7.
34. Chapman GB, Bergus GR, Elstein AS. Order of information affects clinical judgment. *J Behav Decis Making*. 1996;9(3):201–11.
35. Cunnington JP, Turnbull JM, Regehr G, Marriott M, Norman GR. The effect of presentation order in clinical decision making. *Acad Med*. 1997;72(10 Suppl I):S40–S2.
36. Bergus GR, Levin IP, Elstein AS. Presenting risks and benefits to patients—the effect of information order on decision making. *J Gen Intern Med*. 2002;17(8):612–7.
37. Ashton RH, Kennedy J. Eliminating recency with self-review: The case of auditors' 'going concern' judgments. *J Behav Decis Making*. 2002;15(3):221–31.
38. Lim KH, Benbasat I, Ward LM. The role of multimedia in changing first impression bias. *Inform Sys Res*. 2000;11(2):115–36.
39. Rai A, Stubbart C, Paper D. Can executive information systems reinforce biases? *Account Manag Inform Technol*. 1994;4(2):87–106.
40. Shore B. Bias in the development and use of an expert system: Implications for life cycle costs. *Industr Manag Data Sys*. 1996;96(4):18–26.
41. Arnold V, Collier PA, Leech SA, Sutton SG. The effect of experience and complexity on order and recency bias in decision making by professional accountants. *Account Fin*. 2000;40(2):109–34.
42. Arnott D. Cognitive biases and decision support systems development: A design science approach. *Inform Sys J*. 2006;16(1):55–78.
43. Marett K, Adams G, editors. *The role of decision support in alleviating the familiarity bias*. Thirty-ninth Annual Hawaii International Conference on System Sciences (HICSS'06) Track 2; 2006; Hawaii, USA.
44. Clemen RT, Lichtendahl KC, editors. *Debiasing expert overconfidence: A Bayesian calibration model*. Sixth International Conference on Probabilistic Safety Assessment and Management (PSAM6); 2002; San Juan, Puerto Rico, USA.
45. Klayman J, Brown K. Debias the environment instead of the judge: An alternative approach to reducing error in diagnostic (and other) judgment. *Cognition*. 1993;49:97–122.
46. Croskerry P. Cognitive forcing strategies in clinical decision making. *Ann Emerg Med*. 2003;41(1):110–20.
47. Croskerry P. The importance of cognitive errors in diagnosis and strategies to minimize them. *Acad Med*. 2003;78(8):775–80.
48. Jolls C, Sunstein CR. Debiasing through law. *J Legal Stud*. 2006;35:199–241.
49. Roy MC, Lerch FJ. Overcoming ineffective mental representations in base-rate problems. *Inform Sys Res*. 1996;7(2):233–47.
50. Chandra A, Krovi R. Representational congruence and information retrieval: Towards an extended model of cognitive fit. *Decis Supp Sys*. 1999;25(4):271–88.
51. Hearst MA. User interfaces and visualization. In: Baeza-Yates R, Ribeiro-Neto B, editors. *Modern information retrieval*. 1st ed. Boston, MA: Addison-Wesley; 1999. p. 257–324.
52. Coiera E, Walther M, Nguyen K, Lovell NH. Architecture for knowledge-based and federated search of online clinical evidence. *JMIR*. 2005;7(5):e52.
53. Arkes HR. Costs and benefits of judgment errors: Implications for debiasing. *Psychol Bull*. 1991;110(3):486–98.
54. Elting LS, Martin CG, Cantor SB, Rubenstein EB. Influence of data display formats on physician investigators' decisions to stop clinical trials: Prospective trial with repeated measures. *BMJ*. 1999;318(7197):1527–31.
55. Hedwig M, Natter D. Effects of active information processing on the understanding of risk information. *Appl Cog Psychol*. 2005;19(1):123–35.
56. Newell B, Rakow T. The role of experience in decisions from description. *Psychonom Bull Rev*. 2007;14(6):1133–9.
57. Karlsson L, Juslin P, Olsson H. Exemplar-based inference in multi-attribute decision making: Contingent, not automatic, strategy shifts? *Judg Decis Making*. 2008;3(3):244–60.
58. Tubbs RM, Gaeth GJ, Levin IP, Van Osdol LA. Order effects in belief updating with consistent and inconsistent evidence. *J Behav Decis Making*. 1993;6:257–69.
59. Pett MA. *Nonparametric statistics for health care research: Statistics for small samples and unusual distributions*. London: Sage Publications Inc; 1997. p. 79–92.
60. Westbrook JI, Gosling AS, Coiera EW. The impact of an online evidence system on confidence in decision making in a controlled setting. *Med Decis Making*. 2005;25(2):178–85.
61. Purgailis Parker LM, Johnson RE. Does order of presentation affect users' judgment of documents? *J Am Soc Inform Sci*. 1990;41(7):493–4.
62. Huang M, Wang H. The influence of document presentation order and number of documents judged on users' judgments of relevance. *J Am Soc Inform Sci Technol*. 2004;55(11):970–9.
63. Coiera E, Vickland V. Is relevance relevant? User relevance ratings do not predict the impact of internet search engines on decisions. *J Am Med Inform Assoc*. 2008;15:542–5.
64. Sumer HC, Knight PA. Assimilation and contrast effects in performance ratings: Effects of rating the previous performance on rating subsequent performance. *J Appl Psychol*. 1996;81(4):436–42.
65. Smither JW, Reilly RR, Buda R. Effect of prior performance information on ratings of present performance: Contrast versus assimilation revisited. *J Appl Psychol*. 1988;73(3):487–96.
66. Murphy KR, Balzer WK, Lockhart MC, Eisenman EJ. Effects of previous performance on evaluations of present performance. *J Appl Psychol*. 1985;70(1):72–84.
67. Coiera E, Westbrook J, Rogers K. Clinical decision velocity is increased when meta-search filters enhance an evidence retrieval system. *J Am Med Inform Assoc*. 2008;doi:10.1197/jamia.M2765.
68. Friedman C, Gatti G, Franz T, et al. Do physicians know when their diagnoses are correct? Implications for decision support and error reduction. *J Gen Intern Med*. 2005;20(4):334–9.