JAMIA Original Investigations

Research Paper ■

Do People Experience Cognitive Biases while Searching for Information?

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

Abstract **Objective:** To test whether individuals experience cognitive biases whilst searching using information retrieval systems. Biases investigated are *anchoring, order, exposure* and *reinforcement*.

Design: A retrospective analysis and a prospective experiment were conducted to investigate whether cognitive biases affect the way that documentary evidence is interpreted while searching online. The retrospective analysis was conducted on the search and decision behaviors of 75 clinicians (44 doctors, 31 nurses), answering questions for 8 clinical scenarios within 80 minutes in a controlled setting. The prospective study was conducted on 227 undergraduate students, who used the same search engine to answer two of six randomly assigned consumer health questions.

Measurements: Frequencies of correct answers pre- and post- search, and confidence in answers were collected. The impact of reading a document on the final decision was measured by the population likelihood ratio (LR) of the frequency of reading the document and the frequency of obtaining a correct answer. Documents with a LR > 1 were most likely to be associated with a correct answer, and those with a LR < 1 were most likely to be associated with an incorrect answer to a question. Agreement between a subject and the evidence they read was estimated by a *concurrence rate*, which measured the frequency that subjects' answers agreed with the likelihood ratios of a group of documents, normalized for document order, time exposure or reinforcement through repeated access. Serial position curves were plotted for the relationship between subjects' pre-search confidence, document order, the number of times and length of time a document was accessed, and concurrence with post-search answers. Chi-square analyses tested for the presence of biases, and the Kolmogorov-Smirnov test checked for equality of distribution of evidence in the comparison populations.

Results: A person's prior belief (anchoring) has a significant impact on their post-search answer (retrospective: $P < 0.001$; prospective: $P < 0.001$). Documents accessed at different positions in a search session (order effect [retrospective: $P = 0.76$; prospective: $P = 0.026$]), and documents processed for different lengths of time (exposure effect [retrospective: $P = 0.27$; prospective: $P = 0.0081$]) also influenced decision post-search more than expected in the prospective experiment but not in the retrospective analysis. Reinforcement through repeated exposure to a document did not yield statistical differences in decision outcome post-search (retrospective: $P = 0.31$; prospective: $P = 0.81$).

Conclusion: People may experience anchoring, exposure and order biases while searching for information, and these biases may influence the quality of decision making during and after the use of information retrieval systems.

■ **J Am Med Inform Assoc.** 2007;14:599–608. DOI 10.1197/jamia.M2411.

Introduction

Information search plays an important role in supporting the practice of evidence-based medicine[1] and in consumer healthcare decision making.[2] Health-related decisions improve with better access to relevant information, and the Web is increasingly the source of that information.[3] While much research focuses on the design of retrieval methods that identify potentially relevant documents, there has been little examination of the way that retrieved documents then impact decision making.[4,5]

We know that human beings seldom follow a purely rational or normative model in decision making and that we are prone to a series of decision biases.[6] Decision making research has for a long time identified that these biases can have adverse impacts on decision outcomes.[6–9] Medical practitioners display cognitive biases when making clinical decisions[7,8] and interpreting research evidence.[9]

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

This research was supported by an Australian Research Council SPIRT grant and APAI scholarship C00107730, and by Australian Research Council Discovery Grant DP0452359. The authors would like to thank Professor Johanna Westbrook, Dr Sophie Gosling and Miss Nerida Creswick for providing the data in this retrospective analysis; and also Dr David Thomas and Dr Ilse Blignault for their assistance in subject recruitment and the development of the health consumer case scenarios. The authors sincerely thank the reviewers for their insightful and constructive comments which have resulted in substantial improvement to the clarity of this paper.

Correspondence and reprints: Enrico W. Coiera, Centre for Health Informatics, University of New South Wales, UNSW Sydney NSW 2052, Australia; e-mail: <e.coiera@unsw.edu.au>.

Received for review: 2/23/2007; accepted for publication: 6/18/2007.

However, little or no research seems to have examined whether people also experience cognitive biases while searching for information, and whether there are negative consequences from any such biases. Yet, to develop information retrieval systems that actively support health-related decision making, it is necessary to understand the complex process of how people search for and review information when making decisions.[10]

Our own prior analysis of clinician information search used a Bayesian belief revision framework to retrospectively model how documents might influence decisions during and after a search session.[11] The Bayesian model that best predicted the final clinical decision included numerical factors to account for several well known cognitive biases. This study reports a retrospective and a new prospective study that investigate whether people actually experience these cognitive biases while using a Web search engine to answer health-related questions. The biases under investigation are *anchoring, order, exposure* and *reinforcement*.

Background

Biased decisions occur when an individual's cognition is affected by "contextual factors, information structures, previously held attitudes, preferences and moods" (p. 4).[12] Cognitive biases arise because of limitations in human cognitive ability to properly attend to and process all the information that is available (p. 7).[13]

The Anchoring Effect

The anchoring effect, first discussed by Tversky and Kahneman,[14] occurs when a prior belief exerts an influence on the way new information is processed and new beliefs are formed[15] and has been shown to alter clinician and patient judgments.[16–18]

Order Effect

Order effect refers to the way in which the temporal order that information is presented affects a final judgment[19] 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; with recency, impressions are more influenced by later information.[20,21] Studies have shown that medical practitioners arrive at different diagnoses when the same information is presented in a different order,[22–25] and that order influences patients' interpretation of treatment options and information.[26]

Exposure Effect

The degree of exposure to information can affect final judgment. Features in a temporal sequence, such as duration of exposure, the spread of experiences, the partitioning of episodes, and the peak-and-end events have been reported to influence a person's overall impression of the experience.[27]

Reinforcement Effect

Repeated exposure to information may influence the way beliefs are formed. Zajonc[28] found that there was a linear log relationship between the frequency with which a subject was exposed to a stimulus and the subject's enhanced attitude towards the stimulus, regardless of whether the stimulus was a nonsense word, a symbol or a photograph of people.

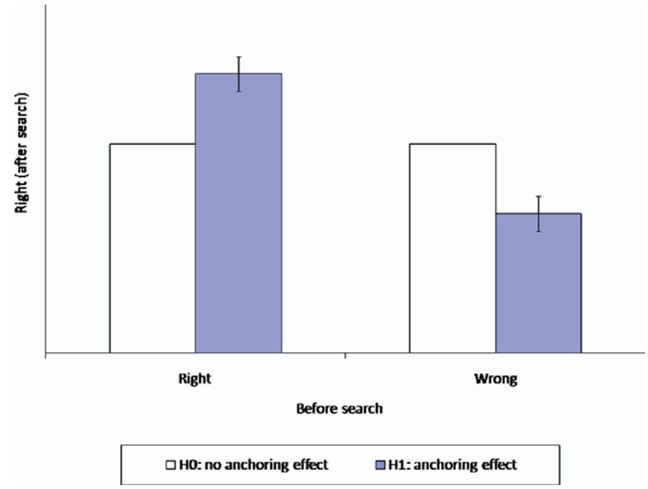

Figure 1. Anchoring effect: the null hypothesis (H0) and the alternative hypothesis (H1).

Hypotheses

We hypothesized that anchoring, order, exposure and reinforcement biases may influence the way individuals interpret documentary evidence retrieved using online search engines.

Anchoring Effect and Confidence in Anchoring Effect Hypotheses

Anchoring effect was tested for with two investigations. The *anchoring effect* investigation assessed whether there was a statistically significant relationship between subjects' pre-search answers and their post-search answers. The *confidence in anchoring effect* investigation looked for a statistically significant relationship between subjects' confidence in their pre-search answer and their tendency to retain a pre-search answer after searching.

If prior beliefs do not affect the way we read a piece of evidence, then we all should arrive at similar conclusions after reading similar impact evidence, irrespective of our past beliefs. The anchoring effect null hypothesis (H0) thus predicts that people are equally likely to answer a question correctly post-search regardless of the correctness of their pre-search answer. H0 expects that the distribution of post-search answers is independent of the distribution of pre-search answers, given access to similar impact evidence by groups with different initial beliefs, i.e., the frequency of right or wrong answers to a given question after search, in the population who answered right before search, should be the same as that of those who provided a wrong answer before search (Figure 1).

However, if prior beliefs do influence how new evidence is interpreted then we should see a reluctance to shift towards new beliefs when new evidence contradicts a prior belief. Thus the alternative hypothesis (H1) states that people who are right pre-search are more likely to be right post-search than subjects who are wrong pre-search. Thus, in the presence of right evidence, we should see fewer pre-search wrong convert to right than expected because they are stubborn to change. Equally, in the presence of wrong evidence we would expect a reluctance to shift by the initially right, to wrong post-search. People might also be

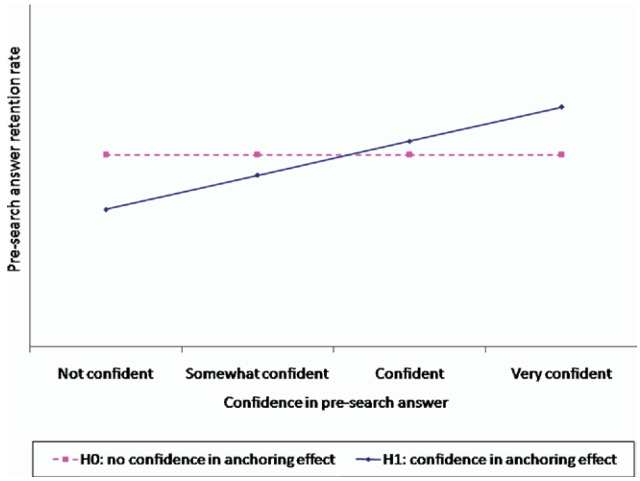

Figure 2. Confidence in anchoring effect: the null hypothesis (H0) and the alternative hypothesis (H1).

more likely to retain a belief in the presence of confirming evidence. For example, subjects may hold a right belief only weakly pre-search, misinterpret right evidence as wrong, and change their view. We should see less of this transition when an anchoring bias is in operation, because the prior belief has a greater impact on the way evidence was interpreted.

For the confidence in anchoring effect (Figure 2) the null hypothesis is that an individual's tendency to change their answer after search is independent of their confidence in the pre-search answer. The alternative hypothesis predicts that people who are more confident in their pre-search answer are more likely to retain their pre-search answer after search than those who are not as confident.

Order Effect Hypothesis

For the order effect (Figure 3) 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. 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. This increase in concurrence would occur because the greater an impact that a piece of evidence has on an individual, the more likely they are to agree with the answer suggested by the evidence.

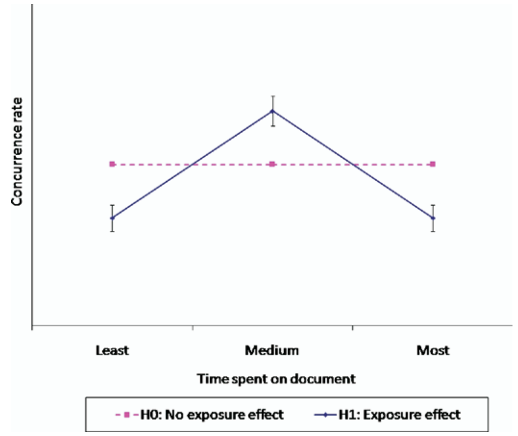

Figure 4. Exposure effect: the null hypothesis (H0) and the alternative hypothesis (H1).

Exposure Effect Hypothesis

For the exposure effect (Figure 4) the null hypothesis predicts that the concurrence between subjects' post-search answer and the answer suggested by a document is not influenced by the amount of time subjects spent on the document. The alternative hypothesis suggests that the amount of time subjects spent on a document influences the concurrence rate.

Reinforcement Effect Hypothesis

For the reinforcement effect (Figure 5) the null hypothesis predicts that the concurrence between subjects' post-search answer and the answer suggested by a document is not influenced by the frequency with which a document was accessed. The alternative hypothesis predicts that an increase in the access frequency of a document will increase the concurrence rate.

Methods

Retrospective Data Analysis

A retrospective analysis was constructed on a dataset of search and decision behaviors of 75 clinicians (44 doctors

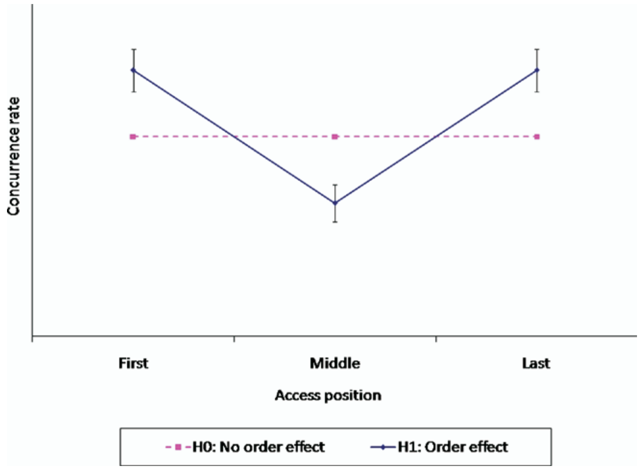

Figure 3. Order effect: the null hypothesis (H0) and the alternative hypothesis (H1).

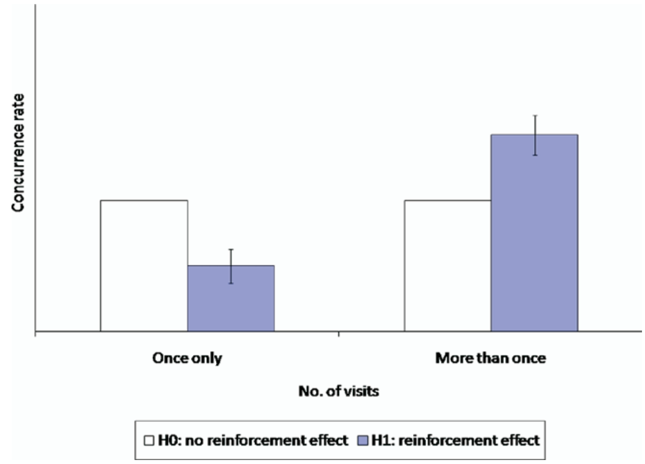

Figure 5. Reinforcement effect: the null hypothesis (H0) and the alternative hypothesis (H1).

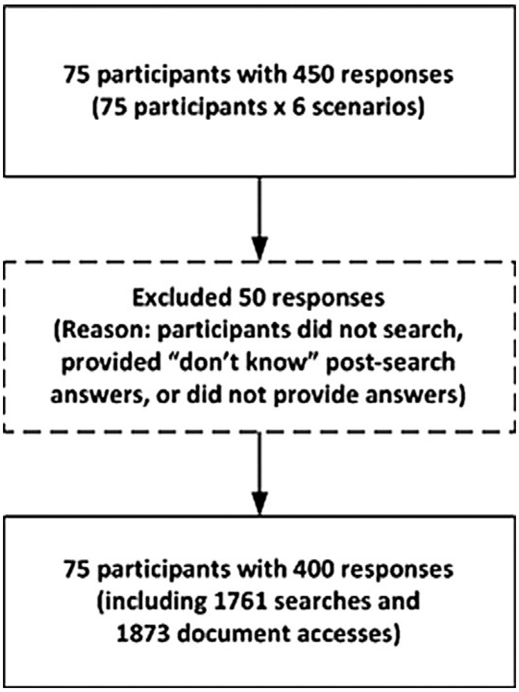

Figure 6. Data exclusion in retrospective analysis.

and 31 clinical nurse consultants), who answered questions for 8 clinical scenarios within 80 minutes in a controlled setting at a university computer laboratory.[4] In that study, scenarios were presented in a random order. After answering a scenario question, each subject was again presented the same scenario and asked to locate documentary evidence to support their answers using a given Internet search engine which provided access to a broad range of resources including PubMed, clinical guidelines and selected journals. Data from six of the eight scenarios in that study for which a correct answer could be identified were included in our retrospective analysis (Table 1).

These 75 subjects recorded their pre- and post-search answers to each question, confidence in these answers and their confidence in the evidence they found. For four questions, there were four answer options: yes, no, conflicting evidence and don't know. Two questions (asthma and diabetes) required users to input a free-text answer. Confidence was measured by a 4 point Likert scale from "very confident" to "not confident." Subjects recorded any change in answer or confidence from their pre-search response and identified which documents influenced their decision. They were asked to work through the scenarios as they would within a clinical situation and not spend more than 10 minutes on any one question.

Prospective Experiment

A convenience sample of 227 non-clinicians was recruited from the undergraduate student population at The University of New South Wales (UNSW). People with Internet access who had previously used an online search engine were recruited by announcements seeking volunteers advertised via student email lists, posters, leaflets, weekly student magazines and research news website on the UNSW campus. Upon completion of the study, subjects were remunerated by being entered into a draw for one of 100 movie tickets.

The pre-/post-search protocol used in the retrospective study was again used in this prospective experiment to allow comparison between the two experiments. Each subject in the prospective experiment answered a set of six questions designed for health consumers and recorded any change in answer or confidence from the pre-search response. The first two questions were answered using the same search engine that was used in the retrospective study and these data are reported here. The remaining four questions were answered using a modified user interface to the same search engine and are excluded from this study because they are not directly comparable interventions.

Questions ranged in difficulty and topic to cover a spectrum of health consumer questions. Each question and the expected correct answer are shown in Table 2. Questions were developed in consultation with a general practitioner and two academics from the School of Public Health and Community Medicine at UNSW. Agreement was reached on the "correct" answer and the location of the best evidence sources for each question. A pilot of three members from the general public tested the questions for interestingness and readability. Two additional pilots of five people each used the system to confirm it was possible to locate documentary evidence required to answer questions correctly. Subjects in both the retrospective and prospective studies were not informed that the purpose of the experiment was to evaluate whether people experience cognitive biases during information searching.

Document Likelihood Ratios

To model the impact an individual document may have had on a subject's decision we calculated the frequency of association between accessing a specific document and a subject making a correct or incorrect decision after accessing that document. A likelihood ratio (LR) was calculated from the ratio of the frequency that accessing a document is associated with a correct answer to the frequency that

Table 1 ▪ Clinical Questions in the Scenarios Presented to Subjects (from Westbrook et al., 2005)

Question (scenario name)	Expected correct answer
Does current evidence support the insertion of tympanostomy tubes in child with normal hearing? (Glue ear)	No, not indicated
What is the best delivery device for inhaled medication to a child during moderate asthma attack? (Asthma)	Spacer (holding chamber)
Is there evidence for the use of nicotine replacement therapy after myocardial infarction? (MI)	No, use is contraindicated
Is there evidence for increased breast and cervical cancer risk after IVF treatment? (IVF)	No evidence of increased risk
Is there evidence for increased risk of SIDS in siblings of baby who died of SIDS? (SIDS)	Yes, there is an increased risk
What is the anaerobic organism(s) associated with osteomyelitis in diabetes? (Diabetes)	*Peptrostreptococcus, Bacteroides*

Table 2 ■ 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. *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?* (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. *Can you catch Hepatitis B from kissing on the cheek?* (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. *Are there different recommendations regarding safe alcohol consumption for males and females?* (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. *Is there an increased risk of a baby dying from SIDS if the mother smokes during pregnancy?* (SIDS)	Yes
5. Breast cancer is one of the most common types of cancer found in women. *Is there an increased chance of developing breast cancer for women who have a family history of breast cancer?* (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. *In Australia, do more men die by committing suicide than women?* (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. *Is there evidence to support the taking of vitamin C supplements to help prevent the common cold?* (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. *Is it likely that we can get AIDS from a mosquito bite?* (AIDS)	No

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

accessing a document is associated with an incorrect answer (Equation 1).

$$Likelihood\ ratio = \frac{P(AccessedDoc|Correct)}{P(AccessedDoc|Incorrect)} = \frac{Sensitivity}{1 - Specificity} \quad (1)$$

The LR is interpreted as a measure of the impact that accessing a document has in influencing a subject towards a specific answer. 'Positive' documents with a LR > 1 are more likely to be associated with a correct answer and 'negative' documents with a LR < 1 are more likely associated with an incorrect answer. A LR around 1 means that a document is equally associated with correct and incorrect answers and probably has no strong influence. In the investigations of order, exposure and reinforcement effect, the answer suggested by a document is represented by the LR of the document.

To calculate the likelihood ratio, the sensitivity and specificity of a document with respect to an answer (correct or incorrect) are calculated. The *sensitivity*, or true positive rate, of a document is the frequency with which the document being accessed correlated with a correct answer being provided post-search (Equation 2). The false negative rate, the frequency with which access of a document correlated with an incorrect answer, was also calculated. The *specificity*, or true negative rate, is one minus the false negative rate (Equation 3).

$$Sensitivity = \frac{No.\ of\ correct\ post\ search\ answers\ where\ document\ was\ accessed}{Total\ no.\ of\ post\ search\ correct\ answers} \quad (2)$$

$$1 - Specificity = \frac{No.\ of\ incorrect\ post\ search\ answers\ where\ document\ was\ accessed}{Total\ no.\ of\ post\ search\ incorrect\ answers} \quad (3)$$

The sensitivity and specificity measures were calculated based upon the frequency a document was accessed for each scenario. Both measures were recalculated for each subject, and specifically excluded the subject's data, to ensure that the likelihood ratio of a document is independent of the subject being analyzed.

Concurrence Rates

Agreement between a subject and the evidence they read was estimated by a *concurrence rate*, which measured the frequency that subjects' answers agreed with the likelihood ratios of a group of documents, normalized for document order, time exposure or reinforcement through repeated access.

For example, to test for the impact of positional biases from the order effect, we calculated the degree of agreement or concurrence between the post-search answers given by subjects and the answer suggested by the documents accessed at different points in a search session, looking to see if there is increased concurrence for documents accessed at some points in a session compared to others. The result is a serial position curve.[29]

To calculate concurrence rates we assembled all the documents accessed by subjects in sessions of length more than one document, where a bias might be detected. To allow sessions of different lengths to be compared, position in a search session was normalized to first, middle and last.

For example, the concurrence rate for documents accessed at the first position in the search session is the number of times a positive document (LR > 1) is read at the first position and the post-search answer is also correct, plus the number of times a negative document (LR < 1) is read at the first position and the post-search answer is also incorrect, divided by the total number of times documents were read at position one in a session (Equation 4).

$$\text{Concurrence rate}_{\text{first}} = \frac{\text{Concurrence}_{\text{first}}}{\text{Access}_{\text{first}}} \tag{4}$$

$\text{Concurrence}_{\text{first}}$
= No. of post-search correct subjects who accessed a positive document at $\text{position}_{\text{first}}$
+ No. of post-search incorrect subjects who accessed a negative document at $\text{position}_{\text{first}}$

$\text{Access}_{\text{first}}$
= Total no. of subjects who accessed a document at $\text{position}_{\text{first}}$

Similar analyses were performed for other positions in a session (i.e., middle and last position), and to test for other biases using the length of exposure to a given document (i.e., normalized to minimum, medium and maximum exposure), the number of times the same document is viewed in a session (i.e., accessed once and more than once) and the confidence in anchoring effect (i.e., not confident, somewhat confident, confident and very confident in pre-search answer).

Quantitative Analyses

Chi-square analysis was conducted for each cognitive bias to test for a statistically significant relationship between documents accessed at different points in a search session and the concurrence between subjects' post-search answers and the answers suggested by documents (i.e., order, exposure and reinforcement effect). It was also used to test whether there was a statistically significant relationship between subjects' pre-search answers and their post-search answers (anchoring effect), as well as between subjects' confidence in the pre-search answer and their tendency to retain the pre-search answer after searching (confidence in anchoring effect).

The null and alternative hypotheses for each bias assume that subjects being compared had equal access to similar impact evidence. If this was not the case then any difference we might detect could be attributed not to any bias but due to the impact of an overrepresentation of right or wrong evidence in our sample unduly influencing the decision outcome one way or another. To test the anchoring hypotheses, subjects who were right pre-search should have had access to equal proportions of evidence supporting a right or a wrong answer as those subjects who were wrong pre-search. Similarly tests for the other biases assumed that the distribution of likelihood ratios amongst accessed documents is uniform across different positions, different exposure levels and different access frequency in a search session. The Kolmogorov-Smirnov (K-S) test for comparing distributions was used to evaluate if distributions of document likelihood ratios differed significantly for any of the document populations being compared.

Results

Data Description

The overall unit of measure in these experiments is a search session, which is "the entire series of queries by a user"[30] to answer one question. Figure 6 describes the data exclusion procedure in the retrospective data analysis. After data exclusion, the retrospective analysis included 75 clinician subjects who made 1761 searches and accessed 1873 documents across the 400 search sessions for six scenarios. On average, clinicians took 405 (standard deviation (SD): 590.8) seconds, made 4.32 (SD: 4.002) searches and accessed 4.65 (SD: 3.670) documents to complete a question in a search session.

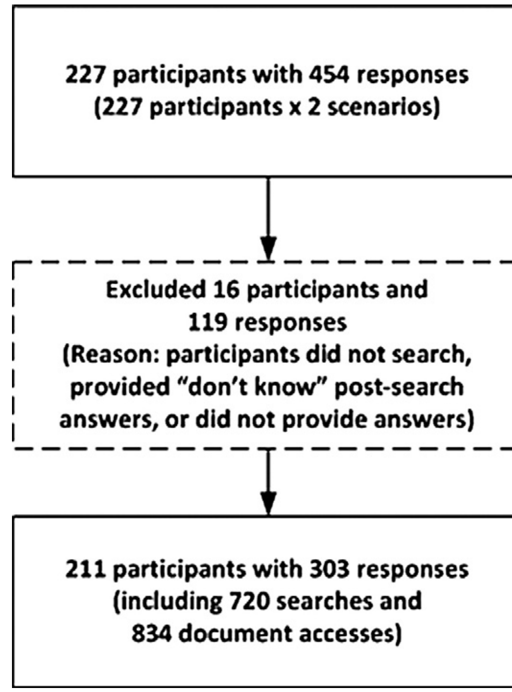

Figure 7. Data exclusion in prospective experiment.

Figure 7 describes the data exclusion procedure in the prospective experiment. After data exclusion, the experiment included 211 health consumer subjects who made 720 searches and accessed 834 documents across 303 search sessions for two questions. On average for all questions, health consumers took 361 seconds (SD: 281.2), made 1.73 (SD: 1.391) searches and accessed 3.25 (SD: 3.067) documents to answer a question in a search session.

Anchoring Effect and Confidence in Anchoring Effect

Anchoring effect data are presented in Table 3 and Figure 8. K-S showed no differences in distribution of evidence in the prospective sample (K-S Z = 1.209, |D| = 0.103, P = 0.108). For the retrospective analysis, using the K-S test, pre-search right and pre-search wrong subjects accessed right and wrong impact documents in unequal proportions (K-S Z = 1.527, |D| = 0.341, P = 0.019). Interestingly, the pre-search

Table 3 ▪ Relationship between Pre-search Answer and Post-search Answer

	After search	
Before search	Right	Wrong
Retrospective analysis		
Right (n=134)	104 (77.6%)	30 (22.4%)
Wrong (n=266)	146 (54.9%)	120 (45.1%)
Prospective experiment		
Right (n=192)	181 (94.3%)	11 (5.7%)
Wrong (n=111)	69 (62.2%)	42 (37.8%)

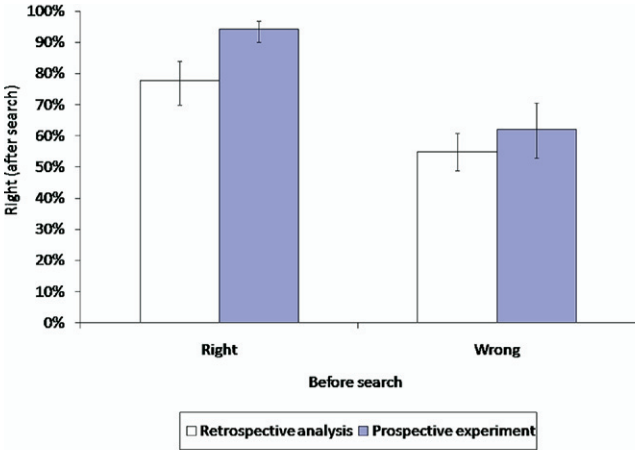

Figure 8. Relationship between pre-search answer and post-search correctness (please refer to Table 3 for frequency numbers).

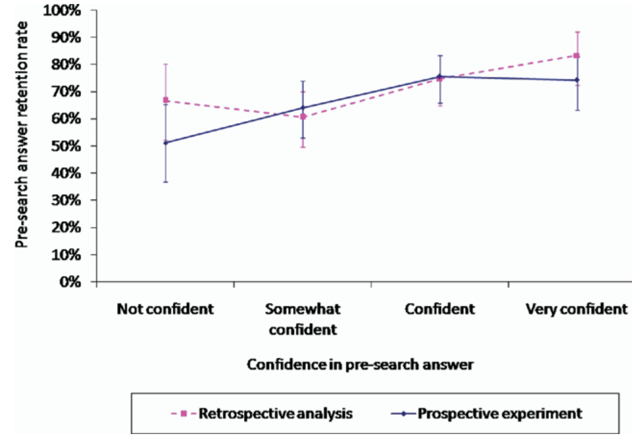

Figure 9. Serial anchor curve: relationship between confidence in pre-search answer and retention of pre-search answer after searching.

wrong subjects accessed more right evidence (right-evidence: 62.5% vs. wrong-evidence: 37.5%) and the pre-search right subjects accessed more wrong evidence (wrong-evidence: 65.4% vs. right-evidence: 34.6%), which was also statistically significant ($\chi^2 = 6.34$, $df = 1$, P = 0.012). This distribution of evidence in the retrospective data therefore has the potential to mask any anchoring bias.

Chi-square analyses on Table 3 show that there was a statistically significant relationship between pre-search answers and post-search answers in both the retrospective analysis and the prospective experiment (retrospective: $\chi^2 = 19.63$, $df = 1$, $P < 0.001$; prospective: $\chi^2 = 50.25$, $df = 1$, $P < 0.001$). Subjects who were right pre-search were more likely to be right post-search than those who were wrong pre-search (also illustrated in Figure 8). Similarly, subjects who were wrong pre-search were more likely to be wrong post-search than those who were right pre-search.

Confidence in anchoring effect data are described in Table 4. In the serial anchor curve (Figure 9), confidence in pre-search answers is plotted against the pre-search answer retention rate (the percentage of subjects that retain their pre-search answer after search). Chi-square analyses conducted on Table 4 show that there was a marginally significant relationship between subjects' confidence in their pre-search answers and their retention of the pre-answer after searching in the retrospective analysis ($\chi^2 = 7.70$, $df = 3$, P = 0.053). However, this relationship was *not* found to be significant in the prospective experiment ($\chi^2 = 2.67$, $df = 3$, P = 0.45).

Order Effect

Order effect data are reported in Table 5; serial position curves are illustrated in Figure 10. Using the K-S test in both the retrospective and prospective analyses, there were no statistically significant differences in the distribution of likelihood ratios between documents accessed at first and middle positions (retrospective: K-S Z = 0.788, $|D| = 0.067$, P = 0.564; prospective K-S Z = 0.466, $|D| = 0.043$, P = 0.982), between first and last positions (retrospective: K-S Z = 1.150, $|D| = 0.117$, P = 0.142; prospective: K-S Z = 0.728, $|D| = 0.076$, P = 0.665), nor between middle and last

Table 4 ■ 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
*Retrospective analysis**		
Not confident (n=33)	22 (66.7%)	11 (33.3%)
Somewhat confident (n=94)	57 (60.6%)	37 (39.4%)
Confident (n=71)	53 (74.6%)	18 (25.4%)
Very confident (n=36)	30 (83.3%)	6 (16.7%)
Prospective experiment†		
Not confident (n=33)	22 (66.7%)	11 (33.3%)
Somewhat confident (n=71)	51 (71.8%)	20 (28.2%)
Confident (n=85)	62 (72.9%)	23 (27.1%)
Very confident (n=106)	84 (79.2%)	22 (20.8%)

*166 responses excluded because subjects responded "don't know" or did not report confidence before search.
†8 responses excluded because subjects responded "don't know" or did not report confidence before search.

Table 5 ■ 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
*Retrospective analysis**		
First (n=255)	192 (75.3%)	63 (24.7%)
Middle (n=470)	342 (72.8%)	128 (27.2%)
Last (n=219)	161 (73.5%)	58 (26.5%)
Prospective experiment†		
First (n=185)	171 (92.4%)	14 (7.6%)
Middle (n=342)	306 (89.5%)	36 (10.5%)
Last (n=185)	155 (83.8%)	30 (16.2%)

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

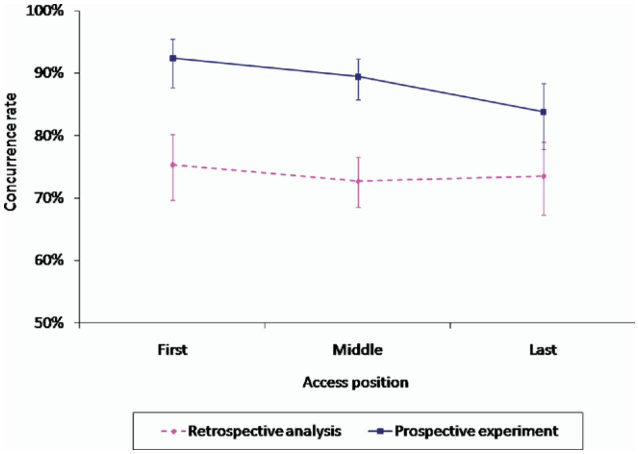

Figure 10. Relationship between document access position and concurrence rate between post-search answer and document-suggested answer.

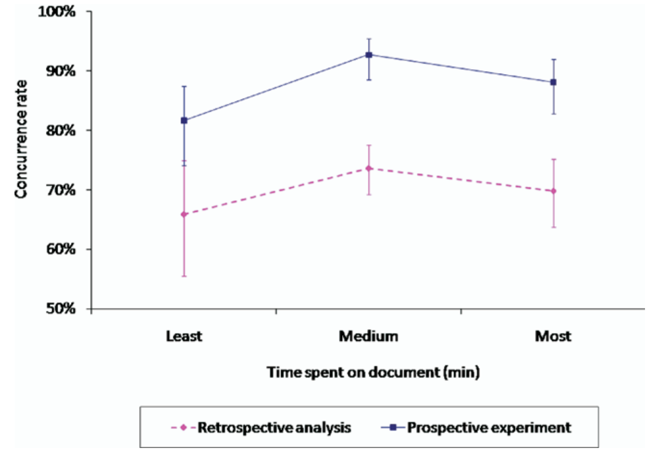

Figure 11. Relationship between document exposure level and concurrence rate between post-search answer and document-suggested answer.

positions (retrospective: K-S Z = 0.913, |D| = 0.082, P = 0.375; prospective: K-S Z = 1.014, |D| = 0.093, P = 0.255).

Chi-square analyses conducted on Table 5 show that there was not a statistically significant relationship between the access position of documents and the concurrence between the post-search answer and the answer suggested by documents in the retrospective analysis (χ^2 = 0.55, *df* = 2, P = 0.76). However, this relationship was statistically significant in the prospective experiment (χ^2 = 7.27, *df* = 2, P = 0.026), where the concurrence rate decreased as the document access position proceeded from first, middle to last (Figure 10).

Exposure Effect

Exposure effect data are reported in Table 6. The serial exposure curves (Figure 11) plot amount of time exposed to a document normalized to the search session (least refers to the document in which the least amount of time was spent in a session, most to the document in which the most amount of time was spent, and medium for all other documents)

Table 6 ■ Relationship between Document Exposure Level and Concurrence between Post-search Answer and Document-suggested Answer

	Concurrence between post-search answer and document?	
Level of exposure	Yes	No
*Retrospective analysis**		
Least (n=88)	58 (65.9%)	30 (34.1%)
Medium (n=417)	307 (73.6%)	110 (26.4%)
Most (n=245)	171 (69.8%)	74 (30.2%)
Prospective experiment†		
Least (n=126)	103 (81.7%)	23 (18.3%)
Medium (n=221)	205 (92.8%)	16 (7.2%)
Most (n=194)	171 (88.1%)	23 (11.9%)

*1123 accessed documents excluded because subjects accessed only one document or a likelihood ratio could not be calculated for the document.

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

against the concurrence rate between subjects' post-search answers and the document-suggested answer.

The K-S test shows that there were no statistically significant differences in the distribution of likelihood ratios between documents accessed at minimum and medium levels of exposure (retrospective: K-S Z = 0.589, |D| = 0.076, P = 0.879; prospective: K-S Z = 0.727, |D| = 0.081, P = 0.665), between minimum and maximum (retrospective: K-S Z = 1.011, |D| = 0.138, P = 0.259; prospective: K-S Z = 0.342, |D| = 0.039, P = 1.000), nor between medium and maximum (retrospective: K-S Z = 0.955, |D| = 0.082, P = 0.322; prospective: K-S Z = 0.632, |D| = 0.062, P = 0.819) in both the retrospective and prospective analyses.

Chi-square analyses conducted on Table 6 show that there was *not* a statistically significant relationship between exposure level of documents and the concurrence between the post-search answer and the answer suggested by documents in the retrospective analysis (χ^2 = 2.61, *df* = 2, P = 0.27). However, this relationship was statistically significant in the prospective experiment (χ^2 = 9.64, *df* = 2, P = 0.0081), where documents processed at most or least exposure in a search session are not as influential on the post-search answer as those that were processed at medium exposure (illustrated in Figure 11).

Reinforcement Effect

Reinforcement effect data are reported in Table 7. The serial reinforcement curves (Figure 12) plot the access frequency of documents in a session in normalized form (once only or more than once) against the concurrence rate between subjects' post-search answers and the document-suggested answer. The K-S test shows that there were no statistically significant differences in the distribution of likelihood ratios between documents accessed once only and documents accessed more than once in the prospective analysis (K-S Z = 0.446, |D| = 0.056, P = 0.989). There was a difference in the distributions for the retrospective analysis (K-S Z = 1.443, |D| = 0.144, P = 0.031). Chi-square analyses conducted on Table 7 show that there was *not* a statistically significant relationship between access frequency of a document and the concurrence between the post-search answer and the answer suggested by the document in both the

Table 7 ■ Relationship between Document Access Frequency and Concurrence between Post-search Answer and Document-suggested Answer

Access frequency	Concurrence between post-search answer and document? Yes	No
*Retrospective analysis**		
Once only (n=441)	306 (69.4%)	135 (30.6%)
More than once (n=156)	115 (73.7%)	41 (26.3%)
Prospective experiment†		
Once only (n=583)	504 (86.4%)	79 (13.6%)
More than once (n=72)	63 (87.5%)	9 (12.5%)

*1276 accessed documents excluded because subjects accessed only one document or a likelihood ratio could not be calculated for the document.
†179 accessed documents excluded because subjects accessed only one document or a likelihood ratio could not be calculated for the document.

retrospective analysis and the prospective experiment (retrospective: $\chi^2 = 1.03$, $df = 1$, $P = 0.31$; prospective: $\chi^2 = 0.061$, $df = 1$, $P = 0.81$) (also illustrated in Figure 12).

Discussion

Results from the prospective experiment support the hypotheses that people experience anchoring, exposure and order biases during information searching. The retrospective data analysis provided support for anchoring, but did not provide evidence for order, exposure and reinforcement.

The differences in these results may be explained in a number of ways. Firstly the retrospective data analysis is post-hoc, i.e. it is not an experiment designed to test for the effects of cognitive biases on information searching and decision making. Secondly, it is possible that each of these biases exerts only a small effect on the decision outcome and that the sample size in the retrospective data analysis was too small to detect them. Finally, the retrospective analysis occurred with clinicians and the prospective analysis on university students who were asked to answer health consumer questions. It may be that there is a novice-expert phenomenon at play here, and that clinician training in some way influences the way information is processed, although the strong evidence for these biases in other aspects of clinical decision making does not support such a notion.

In both the retrospective data analysis and the prospective experiment, there was a statistically significant relationship between subjects' pre-search answers and their post-search answers (anchoring effect), providing the strongest evidence for this effect of all the biases tested for. Subjects who were correct pre-search were more likely to answer correctly post-search than those who were incorrect pre-search. Although the K-S test indicated that there was an unequal distribution of evidence between the pre-search right and wrong groups in the retrospective analysis, this would most likely have masked any bias effect, rather than enhance it. Only the retrospective data analysis showed a marginal statistical relationship between subjects' confidence in their pre-search answers and their tendency to retain that answer after searching (confidence in anchoring effect) suggesting this effect requires further investigation.

Documents accessed at different positions in the search session were found to have different degrees of influence on the post-search answer (order effect) in the prospective experiment. Interestingly the combined primacy and recency effects we hypothesized would generate a u-shaped curve (Figure 3), but with our data, the concurrence rate of a document decreased as the document access position proceeded from first, middle to last. In other words, the degree of influence a document has on a decision decreases as more documents were accessed in a search session, suggesting the primacy effect is dominant.

Documents processed for different durations were found to have different degrees of influence on the post-search answer (exposure effect) in the prospective experiment. Documents which were looked at the most or the least in a search session were not as influential on the post-search answer as those in between.

Documents accessed with different frequencies were not found to have different degrees of influence on the post-search answer in either the retrospective or the prospective analyses. In fact, few subjects accessed documents more than once in both the retrospective study and the prospective experiment. One interpretation of this result is that the reinforcement effect is unlikely to be present during information searching because most people do not access documents more than once.

Limitations

There are several limitations associated with our analysis:

- *Measures:* Core to our analysis is the measure of concurrence between the decision made by a subject, and the expected impact a document has in creating a belief, measured by the document's likelihood ratio. It may be the case that other factors beyond biases affect concurrence rates. Document likelihood ratios are an aggregate measure of document influence over a population, and we tested for biases by comparing the impact of different document sets via their pooled LR values, rather than by comparing the impact of individually identifiable documents, which would have required a different and tightly controlled laboratory test, rather than the more natural experiment we sought to create. It may thus be the case that document

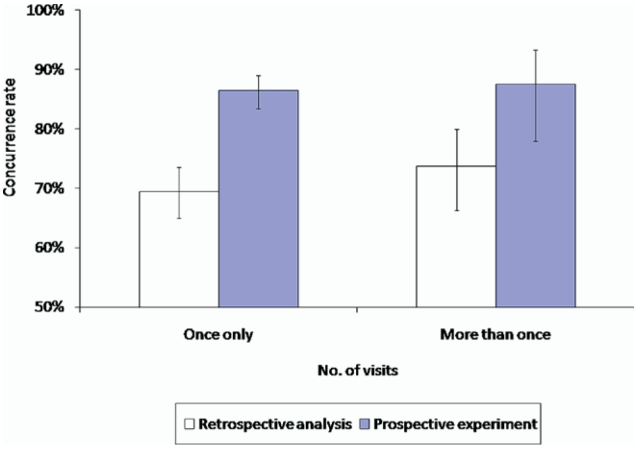

Figure 12. Relationship between document access frequency and concurrence rate between post-search answer and document-suggested answer.

likelihood ratios and concurrence rates are not ideal measures of the underlying phenomena we wish them to model. For the exposure bias, we only looked at the impact of length of exposure, but this bias, like the others, may be manifest in other ways than those tested for.

- *Interaction of biases:* This investigation studies one cognitive bias at a time. It is unclear how biases interact and the *collective* impact they may have on information searching and decision making.

Conclusion

This research is possibly the first study that looks at the impact of cognitive biases on information searching and decision making. It provides evidence that people can experience cognitive biases while searching for information. A person's prior belief (anchoring effect), the order in which documents were accessed (order effect) and the amount of time people spend on documents (exposure effect) are factors found to have influenced the way questions were answered post-search.

However, there is a lack of benchmark literature in this area. More experiments are needed to test the findings in this study before generalizations can be made on how widespread and significant these biases are. Other cognitive biases, such as confirmation bias and conservatism bias[15,31] should also be investigated for their prevalence in information searching. In addition, further investigations could examine whether biases experienced during information searching can be moderated to improve the impact of retrieved evidence on the quality of decision making, and may eventually lead to better designs for search systems.

References ■

1. Hersh WR. Evidence-based medicine and the internet. ACP J Club 1996;125:A14–6.
2. Eysenbach G, Jadad A. Evidence-based patient choice and consumer informatics in the internet age. J Med Internet Res. 2001;3:e19.
3. Morrison JB, Pirolli P, Card SK. A taxonomic analysis of what world wide web activities significantly impact people's decisions and actions. Conference on Human Factors in Computing System. New York: ACM Press; 2001.
4. Westbrook JI, Coiera EW, Gosling AS. Do online information retrieval systems help experienced clinicians answer clinical questions? J Am Med Inform Assoc 2005;12:315–21.
5. Hersh WR. Ubiquitous but unfinished: On-line information retrieval systems. Med Decis Making 2005;25:147–8.
6. Kahneman D, Slovic P, Tversky A. Judgment under uncertainty: Heuristics and biases. New York: Cambridge University Press; 1982.
7. Elstein AS. Heuristics and biases: Selected errors in clinical reasoning. Acad Med 1999;74:791–4.
8. Croskerry P. Achieving quality in clinical decision making: Cognitive strategies and detection of bias. Acad Emerg Med 2002;9:1184–204.
9. Kaptchuk TJ. Effect of interpretive bias on research evidence. BMJ 2003;326:1453–5.
10. Spink A, Cole C. New directions in cognitive information retrieval. 1st ed. London: Springer; 2005.
11. Lau AYS, Coiera EW. A bayesian model that predicts the impact of web searching on decision making J Am Soc Inform Sci Technol 2006;57:873–80.
12. 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.
13. Kruglanski AW, Ajzen I. Bias and error in human judgment. Eur J Soc Psychol. 1983;13:1–44.
14. Tversky A, Kahneman D. Judgment under uncertainty: Heuristics and biases. Science. 1974;185:1124–31.
15. Wickens CD, Hollands JG. Decision making. Engineering psychology and human performance. 3rd ed. Upper Saddle River, NJ: Prentice Hall; 2000. pp. 293-336.
16. 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;2:203–11.
17. Berner E, Maisiak right, Heudebert G, Young K. Clinician performance and prominence of diagnoses displayed by a clinical diagnostic decision support system AMIA Ann Symp Proc 2003;76–80.
18. Friedlander ML, SJ. S. Anchoring and publicity effects in clinical judgment. J Clin Psychol 1983;39:637–43.
19. Wang H, Zhang J, Johnson TR. Human belief revision and the order effect. Twenty-second Annual Conference of the Cognitive Science Society. Hillsdale, NJ; 2000.
20. 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.
21. Luchins AS. Primacy-recency in impression formation. In: Hovland CI, Mandell E, Campbell E, Brock T, Luchins AS, Cohen A, et al., eds. The order of presentation in persuasion. New Haven, CT: Yale University Press; 1957. pp. 33–61.
22. 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;27:314–20.
23. Bergus GR, Chapman GB, Levy BT, Ely JW, Oppliger RA. Clinical diagnosis and the order of information. Med Decis Making 1998;18:412–7.
24. Chapman GB, Bergus GR, Elstein AS. Order of information affects clinical judgment. J Behav Decis Making 1996;9:201–11.
25. Cunnington JP, Turnbull JM, Regehr G, Marriott M, Norman GR. The effect of presentation order in clinical decision making. Acad Med 1997;72:S40–S2.
26. 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;17:612–7.
27. Varey CA, Kahneman D. Experiences extended across time: Evaluation of moments and episodes. J Behav Decis Making 1992;5.
28. Zajonc RB. Attitudinal effects of mere exposure. J Pers Soc Psychol (Monograph Suppl) 1968;9:1–27.
29. 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.
30. Jansen BJ, Spink A. An analysis of web documents retrieved and viewed. Fourth International Conference on Internet Computing. Las Vegas, Nevada; 2003. p. 65-9.
31. Elstein A S. Clinical problem solving and diagnostic decision-making: A selective review of the cognitive literature. BMJ 2002;324:729–32.