# Cognitive Biases in Search

## A review and reflection of cognitive biases in Information Retrieval

Leif Azzopardi
Leifos@acm.org
Glasgow, UK

## ABSTRACT

People are susceptible to an array of cognitive biases, which can result in systematic errors and deviations from rational decision making. Over the past decade, an increasing amount of attention has been paid towards investigating how cognitive biases influence information seeking and retrieval behaviours and outcomes. In particular, how such biases may negatively affect decisions because, for example, searchers may seek confirmatory but incorrect information or anchor on an initial search result even if its incorrect. In this perspectives paper, we aim to: *(1)* bring together and catalogue the emerging work on cognitive biases in the field of Information Retrieval; and *(2)* provide a critical review and reflection on these studies and subsequent findings. During our analysis we report on over thirty studies, that empirically examined cognitive biases in search, providing over forty key findings related to different domains (e.g. health, web, socio-political) and different parts of the search process (e.g. querying, assessing, judging, etc.). Our reflection highlights the importance of this research area, and critically discusses the limitations, difficulties and challenges when investigating this phenomena along with presenting open questions and future directions in researching the impact — both positive and negative — of cognitive biases in Information Retrieval.

## KEYWORDS

Cognitive Bias, Heuristics, Search, Information Retrieval

**ACM Reference Format:**
Leif Azzopardi. 2021. Cognitive Biases in Search: A review and reflection of cognitive biases in Information Retrieval. In *Proceedings of the 2021 ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR '21), March 14–19, 2021, Canberra, ACT, Australia.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3406522.3446023

## 1 INTRODUCTION

*Information Seeking and Retrieval* (ISR) is a process of searching, discovering and finding relevant, useful, and credible information [30]. Many factors impact upon how people undertake this process, shape their search behaviours, and effect their search performances. Over the years, researchers have investigated user characteristics (e.g., expertise, background, topic knowledge, cognitive abilities, etc.), system functionalities (e.g., interface, presentation, quality, etc.), task attributes (e.g., task difficulty, complexity, topicality, etc.) [36, 40] and many other factors. More recently, there has been growing interest in exploring how cognitive biases influence people's ISR and how they impact upon people's information processing, knowledge acquisition and decision making. Of particular concern to the field is whether the influence of cognitive biases may be exacerbated due to the instantaneous access to unprecedented volumes of information, and exploited by search engines and content creators, deliberately or inadvertently [7, 12**?** ]. Concerns have also been raised over whether cognitive biases may also interact with search engine biases, algorithmic biases and content biases [5, 78], or lead to such system sided biases [13, 31] creating a vicious cycle where: "*bias begets bias*" [7]. Taken together, these system- and user-sided biases may compound together, amplifying the effects both positively and negatively [45]. And, so with an increasing amount of the population turning to search and recommender systems to access, find and consume information in order to make important life decisions, regarding for example, medical, political, social, personal and financial choices, investigating the influence and impact of cognitive biases with ISR is of economic and societal importance.

In this paper, we aim to bring together the research on cognitive biases in ISR, cataloguing the main cognitive biases that have been observed in ISR studies — and categorising these studies in terms of their search domain and which part of the search process the cognitive biases manifest. Then, in our discussion, we critically reflect upon this prior work and consider, *"whether we as researchers are suffering from the Observer-Expectancy Effect?"*, while detailing the difficulties, limitations and challenges in studying the influence and impact of cognitive biases in search.

## 2 COGNITIVE BIASES AND HEURISTICS

A cognitive bias is a systematic pattern of deviations in thinking which may lead to errors in judgements and decision-making [74, 75]. These deviations often refer to the difference from what is normatively expected given rational decision making models (e.g., *Probability Theory*, *Expected Utility Theory*, etc.) [74, 75].

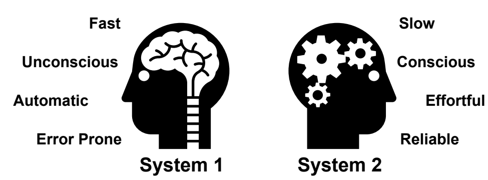


**Figure 1: Thinking, Slow and Fast [32]: Cognitive biases [74], or simple heuristics that make us smart? [73]**


*CHIIR '21, March 14–19, 2021, Canberra, ACT, Australia*

ACM ISBN 978-1-4503-8055-3/21/03…$15.00
https://doi.org/10.1145/3406522.3446023

According to Kahneman [32], cognitive biases arise due to the tension between two conceptual systems within people's brains: *System 1* being automatic, fast and intuitive; and *System 2* being conscious, slow and analytical (see Fig. 1). When system *1* dominates thinking, it can lead to faster decisions, but they can be error prone (see Fig. 2). This is because the biases (and heuristics applied to *think fast*) affect the way in which people perceive and process new information, especially if the information is counter-intuitive, conflicting or induces uncertainty [74]. When system *2* thinking is engaged, it tends to be more reliable but requires more cognitive effort and slows down the decision making process. While it is typically assumed that cognitive biases have a negative impact leading to poorer decisions [74], this is not necessarily always the case, all of the time. Instead cognitive biases may simplify decision-making by reducing the amount of information and uncertainty that has to be processed [32]. And, thus act as mechanisms that enables fast and frugal decision making that makes people smart [73]. For example, people can develop information-processing shortcuts that lead to more effective actions in a given context, or enable faster decisions when timeliness is more valuable than accuracy. So rather than being universally detrimental to decision making, it has been shown that these *fast and frugal heuristics* that people employ result in good decisions most of the time [24, 73]. This is because, they argue, that people seeks to do their best, with their available information processing capabilities, given the constraints of the environments (e.g., bounded or computational rationality) [25, 48, 70]. Essentially, cognitive biases can have both positive and negative impacts – though most of the literature within ISR has focused on their negative impacts.

Since the seminal work by Tversky and Kahneman [74] over 180 different biases and effects have been identified [10]. These have been broadly categorised into four high level groupings depending on: *(i)* the amount of information available/presented; *(ii)* the lack of meaning associated with the information; *(iii)* the need to act fast; and *(iv)* what information is remembered or recalled [10]. Below, we describe these four categories, and within these categories, we describe the subset of cognitive biases that have been empirically examined within ISR studies (see Table 1).

A bat and a ball cost 110 cents in total.
The bat costs 100 cents more than the ball.
How much does the ball cost?

**Figure 2: A cognitive reflection test question: What's the first answer that came to your mind? Was it correct[1]?**

## 2.1 Too Much Information

Being overloaded by the amount of information accessible and available may lead to: *(i)* noticing things already primed in memory or that are repeated often (e.g., *Availability Bias*); *(ii)* noticing when something has changed or how it is presented (e.g., *Framing Effects*); *(iii)* drawn to details that support existing beliefs (e.g., *Anchoring Bias*, *Confirmation Bias*); *(iv)* noticing visually striking things; and *(v)* noticing flaws in others more easily than ourselves.

[1]The answer is 5 cents. In [55], only 52% of participants were correct.

**Availability Bias** leads people to overestimate the likelihood of an answer or stance based on how easily it can be retrieved and recalled [74]. Within the context of search, this may mean searchers are more susceptible to system and content biases which promotes content of a particular stance [78], and further compounded because of the *Google Effect*, where people tend to rely on retrieving information via search engines, rather than remembering it [72].

**Framing Effects** occur when people make different decisions given the same information because of how the information has been presented. The classic experiment on framing asked participants which treatment to choose—where the options are presented in terms of the lives saved, or the resulting deaths [74]. In search, framing the presentation of results may also lead the searcher to different perceptions about the topic [53, 54].

**Anchoring Bias** stems from people's tendencies to focus too much on the first piece of information learnt, or observed (even if that information is not relevant or correct) [74]. Anchoring effects may be due to short term anchoring (e.g., an initial result presented may colour the person's opinion on the topic [69]), or stem from entrenched prior beliefs that influence how new information presented during the course of searching is processed and used to make a decision [46].

**Confirmation Bias** stems from people's tendency to prefer confirmatory information, where they will discount information that does not conform to their existing beliefs [51]. When querying, this may manifest as people employing positive test strategies where they try to find information that supports their hypotheses. While, when assessing they may actively dismiss or disregard information that contradicts their hypotheses [42, 76].

## 2.2 Not Enough Meaning

Experiencing certain events may lead to, for example, *(i)* finding patterns in sparse data ; *(ii)* stereotyping or generalising based on prior histories (e.g., *Authority/Trust Bias*, *Bandwagon Effects*); *(iii)* attributing positive attributes to familiar things (e.g., *Reinforcement Effects*, *Exposure Effects*); *(iv)* simplifying probabilities and numbers so they are easier to process; *(v)* assuming that we know what other people are thinking (e.g., *Curse of Knowledge*); and *(vi)* projecting our current mindset onto the past and future (e.g., *Projection Bias*).

**Bandwagon Effects** occur when people take on a similar opinion or point of view because other people voice that opinion or point of view. This is also referred to as *Group Think* or *Herd Behaviour*. For example, the inclusion of popularity and other social indicators, may lead to searchers taking a query suggestion [38], adopting a similar political stance [27], or making a similar judgement regardless of its integrity or accuracy [17].

**Reinforcement Effects** are related to *Bandwagon Effects*, and occur when a stimulus is repeatedly shown to a person, they develop a more positive attitude towards that stimulus. So, if a search result page returns many results purporting a similar stance, this may sway voter opinions [19] or lead searchers to conclude that an ineffective treatment is helpful [46].

**Exposure Effects** are similar to reinforcement effects, but with respect to the amount of time and number of times people spend engaging with the stimulus. So the more time (and times) a person spends engaging with information of a particular stance, viewpoint, etc., and the more exposures they have spread over a period of time, it will tend to lead to a more positive impression of that stance, viewpoint, etc. [19, 46, 78].

## 2.3 Needing to Act Fast

Making decisions under time pressure (or very quickly) may lead to biases in decision making, involving: *(i)* favouring simple options with complete information over complex, ambiguous options (e.g., *Ambiguity Effects*, *Less is More*); *(ii)* avoiding mistakes by preserving autonomy, group status and avoiding irreversible decisions (e.g., *Decoy Effects*, *Status Quo Bias*); *(iii)* getting things done, by focusing on completing tasks where time and effort has been invested (e.g., *Sunken Cost Fallacy*); *(iv)* staying focused by favouring immediate, relatable and available things (e.g., *Novelty Bias*, *Recency Bias*); and *(v)* being confident to act and feeling the actions are important (e.g., *Dunning-Kruger Effect*, *Overconfidence Bias*).

**Ambiguity Effects** occur when several options are presented, people tend to avoid options in which there is high uncertainty in the outcome, even if it is favourable [20]. In search, this may manifest in the selection of result items from known, trusted sources that provide reliable information, rather than selecting items from unknown sources for which there is high uncertainty [34].

**Less is More Effects** arise due to the paradox of choice [68]. When many options are presented, people find it harder to make comparisons, and often will not make any decisions or be less satisfied in the decision that they make —because there are so many options available. Reducing the number of results presented may lead to people being more satisfied in their selection [56].

**Decoy Effects** occur when people's preferences for options $A$ or $B$ changes in favour of option $B$, when a third option $C$ is introduced, where option $C$ is inferior to option $B$. Option $C$ is the decoy and can be clearly distinguished as inferior when compared to $B$, making the comparison between the two cognitively less taxing. However, choosing between $A$ and $B$ is more cognitively taxing, as $A$ maybe be better in one respect, while $B$ in another. Thus, people are more likely to favour $B$ when $C$ is included. This is also referred to as the *Asymmetric Dominance Effect* [28]. In the context of selecting between result items (where one is relevant, and one is not-relevant), the introduction of another non-relevant item which is similar but dominated by the other non-relevant item may deceive searchers into selecting a non-relevant item [17].

**Dunning-Kruger Effect** arises when people (who are generally less competent) overestimate their capabilities in performing a task – and stems for their inability to recognize their lack of capability [16]. When searching, such searchers may then overestimate their ability to find or identity relevant items, which may result in judging non-relevant items as relevant [22].

## 2.4 What to Remember

The need to be selective in retaining information from all that is available may lead to: *(i)* discarding specifics for generalities (e.g., *Fading Affect Bias*, *Negativity Bias*); *(ii)* reducing events and lists to their key elements (e.g., *Recency and Primacy Effects*, *Peak-End Rule*, *Misinformation Effect*); *(iii)* storing memories depending on how they are experienced; and *(iv)* editing and reinforcing memories after the fact.

**Priming Effects** occur when a person's exposure to a stimulus subconsciously influences their response to subsequent stimuli. These stimuli are often related to words or images that people see during the search session. For example, the images and results returned in response to a query may prime the search, activating particular mental concepts that influences their subsequent perceptions – which may lead to very unsavoury, offensive or inappropriate associations being made and propagated (see [52]).

**Order effects** refer to effects based on the order in which information is presented and processed and how that ordering may influence the final decision [74]. These effects can be based on [14]:

- **Primacy** where a person's decision is influenced by the initial information presented in a given sequence/list; and
- **Recency** where a person's decision is influenced by the latter information presented in a given sequence/list.

In search, primacy order effects may mean searchers are given more weight to information presented earlier in the ranked lists relative to information given further down—unless they examine lots of result items, in which case the recency order effects may have a greater impact on their decisions [8, 46].

**Peak End Rule** is a cognitive bias that influences how people recall experiences, where highly charged positive or negative moments during that experience are considered the *"peaks"*—and the final moments of the experience are considered the *"end"*—are weighted more heavily in their assessment. For example, the user's search satisfaction may be highly influenced by these peak and end moments experienced during their search session [49].

The above list of biases outlined is *not exhaustive*, and only represents the majority of cognitive biases that have been studied within ISR contexts specifically (see Table 1). It should be noted that many other cognitive biases, heuristics and effects are also likely to influence ISR behaviours (such as *Attentional Bias*, *Negativity Bias*, *Information Bias*, *Choice-Support Bias*, *Judgment Bias*, etc.). However, we do not have space to review these all here. Instead, we refer the reader to the work by Gluck [26] and the work by Behimehr and Jamali [9] where they have both performed qualitative studies identifying other possible biases that could arise during ISR. In practice, however, given the vast array of cognitive biases, it is often difficult to tease out exactly which cognitive bias influences or shapes a particular decision. More likely, as many interactions are being performed during the search process, a series of heuristics and mental shortcuts will be invoked, leading to a combination of different cognitive biases having an influence [9]. In ISR, it is largely unknown whether: (*i*) certain cognitive biases dominate other cognitive biases, (*ii*) if combinations of cognitive bias compound together leading to greater errors in thinking [45], (*iii*) if certain combinations of cognitive biases may wash out any negative or positive effects, and (*iv*) whether certain cognitive biases, or combinations of, could be invoked to positively influence the searcher to make better decisions [8, 18, 47, 81]. In the literature, however, most research has sought to identify cognitive biases and their negative impacts.

**Table 1: A breakdown of ISR papers investigating different cognitive biases across domains and different parts of the search process.**

| | Cognitive Biases | Domains | | | Search Process | | | |
|---|---|---|---|---|---|---|---|---|
| | | Health | Political | Web | Querying | Examining | Judging | Sat. |
| Too Much Information | Confirmation Bias | [23] [41] [61] [63] [76] [77] [83] | [33] [44] [43] | [62] | [41] [62] [63] | [76] [77] | [23] [61] [43] [33] [44] | |
| | Anchoring | [46] [61] | [53] [54] | [17] [69] | | [53] [54] | [17] [46] [61] [69] | |
| | Availability | [23] [61] [79] | [53] [54] | | | [53] [54] | [23] [61] | |
| | Framing Effects | | [53] [54] | | | [53] [54] | | |
| No Meaning | Bandwagon Effects | [18] [23] [27] | [17] | [11] [38] | [38] | [11] | [17] [18] [27] [23] | |
| | Exposure Effects | [23] [46] [61] | [43] [19] | | | | [19] [23] [43] [46] [61] | |
| | Reinforcement Effects | [46] | [43] [19] | | | | [19] [43] [46] | |
| Act Fast | Decoy Effects | | | [17] | | | [17] | |
| | Ambiguity Effects | | [43] [33] | [17] [29] [34] | | [33] | [17] [29] [34] [43] | |
| | Less is More | | | [56] | | | | [56] |
| | Dunning-Kruger Effect | | | [22] | | | [22] | |
| Remember | Priming Effect | | [53] [54] | [39] [62] [67] [81] | [62] [81] | | [53] [54] [67] | [39] |
| | Order Effects | [8] [46] [61] [1] | [19] | [11] [35] [50] [80] | | [11] [35] [50] [80] [1] | [8] [19] [46] [61] | |
| | Peak End Rule | | | [49] | | | | [49] |

## 3 OVERVIEW OF COGNITIVE BIASES IN IS&R

To seed the discussion and analysis presented in this perspective's paper, we performed a literature review seeking out works in the field of ISR which had performed studies related to cognitive biases specifically, and excluded studies related to algorithmic biases[2]. We identified thirty-five papers that empirically explored the influence of cognitive biases where they test or investigate specific cognitive biases quantitatively, or using a mixed design, and two papers that performed qualitative studies that identify potential cognitive biases that may manifest during the ISR process [9, 26].

Table 1 reports the papers presenting empirical studies broken down by the different cognitive biases presented in Section 2, grouped by the domain: (*i*) Health and Medical, (*ii*) Socio-Political, and (*iii*) Web, and by different parts of the search process: (*i*) querying, (*ii*) examining result lists, (*iii*) judging result items, and (*iv*) assessing search satisfaction. Note that our groupings are not mutually exclusive, and so papers can appear in multiple categories. Below, we present an overview of the key findings from studies in the Health and Socio-Political domains, specifically, and then describe the key findings with respect to the different parts of the search process (as these are mainly web related, these have been combined together rather than repeated).

### 3.1 Cognitive Biases in Health Search

Utilizing search engines to find health and medical information represents a critical domain, as what, where and when the information is presented [46, 47, 78, 79], and how it combines with the searcher's prior beliefs [61, 76] can have a significant impact on health decisions and outcomes [66, 83]. For example, White and Horvitz [79] found that a disproportionate amount of results are returned that overly exaggerate the seriousness of benign conditions (*Availability Bias*). Coupled with people's difficulty in understanding base rates (*Base Rates Fallacy*), this can lead searchers to overestimate the likelihood of them having a particular condition (referred to as *Cyberchrondia*). Many of the studies performed in this domain have used search tasks for which there are clinical answers given Cochrane systematic reviews, such as "*does cinnamon help diabetes?*" (unhelpful), "*Does caffeine help asthma?*" (helpful), etc.. By using such search tasks with known clincial answers, researchers have an objective way to evaluate the accuracy of people's beliefs, the quality of their selections, and veracity of their answers. Researchers have found that:

[2]Note that our review was not a systematic review, and so is not exhaustive, but we believe that it does contain most of the relevant works performed in the area to date.

**HM1** Searchers tend to pose positively framed queries e.g. "*does Aloe Vera cure cancer?*" which suggests a form of *Projection Bias*, where searcher queries may be indicative of their prior belief's, i.e. that Aloe Vera does cure cancer [61, 76].

**HM2** Searchers tend to select results that confirm their beliefs expressed in the query (*Confirmation Bias*) [77]. However, health web content is often biased toward positively framed answers e.g. documents promoting Aloe Vera and its benefits for cancer treatment (*Content Bias* [76, 77]).

**HM3** Searchers tend to anchor on the first result, regardless of whether it is positive (saying the treatment does help) or negative (saying that the treatment does not help), and was more influential on their final answer in [46, 61] (i.e. *primacy effects*), but the converse was found in [1] (i.e. *recency effects*).

**HM4** Searchers appear to value subsequent results less and less, and so latter results have a diminishing influence on their final answer, such that the last result examined had little or the least influence on their final answer in [46], but the converse was found in [1].

**HM5** However, searchers who encountered many correct (incorrect) answers would be more likely to give a correct (incorrect) answer (suggesting *Availability Bias* and/or *Exposure Effects*) [1, 23].

**HM6** Taken together HM1-5 suggest that it is not clear whether a person's existing prior belief, the initial anchor, and/or subsequent results viewed will influence their decision.

**HM7** When searchers invested more effort in the search task, by issuing more queries, and inspecting more items, the accuracy of their answers improved [61].

**HM8** The more domain knowledge searchers had, prior to searching, the more accurate their answers, as they could articulate their queries more precisely, and better understand medical jargon [41]. But, imprecise representations of medical conditions lead searchers to consider more irrelevant information, as well as seeking out information that confirmed their initial belief (*Confirmation Bias*).

**HM9** It was hypothesised that searchers with more domain knowledge can better filter out irrelevant information (*Selective Exposure*), and that once they encounter the first credible and reasonable answer that they will stop searching (*Satisfice*) [41].

Finally, a number of other studies have examined whether it is possible to mitigate these observed biases, or even use biases to nudge people into making better decisions in the health context [8, 18, 47]. Following on from their study in [46], Lau and Coiera [47] created interfaces to help mitigate *Order Effects* and *Anchoring Bias*. They found that *Order Effects* could be mitigated to some extent, but not the *Anchoring Bias* (**HM10**). Bansback et al. [8] exploited *Order Effects* when presenting information to patients, which led to more informed choices about treatments (**HM11**). While, more recently, Elsweiler et al. [18] tried to nudge participants towards healthy recipes by manipulating the attractiveness of the recipes presented (using *Attractiveness Bias*) which led to participant's selecting lower fat recipes (62% of the time, given two choices) (**HM12**).

## 3.2 Cognitive Biases in Socio-Political Search

Deciding on what party to support, whether to become vegan, or deciding to legalize cannabis, represent just a few examples of different socio-political decisions that people may turn to search engines to help inform their opinions [44, 53]. Researchers have been concerned that search engines may be influencing people's opinions, either by presenting confirmatory information reinforcing people's existing beliefs [33, 43], or by presenting information to sway their decisions through exposure effects (dubbed the *Search Engine Manipulation Effect* (SEME) [19]). Similar to health related studies, researchers typically asked participant's their prior beliefs regarding which way they would vote or their opinion on a controversial topic (such as "gun control", "abortion legalisation", etc.). Then after interaction with a search system or result list, they asked them to vote or rate their position again, so that shifts in attitude could be measured [43]. From these socio-policitial studies, researchers have found that:

**SP1** Searchers viewed articles for significantly less time, when the articles were not consistent with their prior beliefs on the topic. While, they spent longer on articles that were consistent [43] (suggesting *Information Avoidance Bias* and *Confirmation Bias*) .

**SP2** Searchers who spent more time viewing articles consistent with their prior beliefs, strengthen their beliefs on the topic [43] (suggesting *Exposure Effects and Reinforcement Effects*).

**SP3** Searchers spent longer viewing articles that were more credible [43, 53, 54] (suggesting *Authority* and *Ambiguity Biases*).

**SP4** While, it was hypothesised that results inconsistent with prior beliefs would be rated as less credible due to potential *Confirmation Bias*, it was found that the prior beliefs of searchers had little influence on their judgments of the credibility of the results [33].

**SP5** Searchers rated articles as more useful if they were easier to read and understand, which they attribute to *Availability Bias*, and searchers engaged less with more complex articles that required more effort to process [53, 54].

**SP6** When presented with a list of results, *Order Effects* and *Exposure Effects* could sway non-voters[3] towards favouring a particular political candidate by roughly 20%, depending on how many results and at what rank the favoured candidate's results were positioned. However, despite non-voters, evaluating and selecting one candidate before examining the result list, no anchoring effects were observed [19].

## 3.3 Cognitive Biases when Querying

In addition to the works previously mentioned examining Confirmation Bias when querying in the Health domain, other studies have examined how searchers may be primed by the querying functionality, either via query auto-complete or query and question suggestions, and how popularity information regarding the suggestions may influence search behaviour. Researchers have found:

**Q1** Searchers selection of query suggestions was not influenced by the displayed popularity of the query suggestion (suggesting that participants were not jumping on the bandwagon) [38].

**Q2** Instead, searchers selected query suggestions that were perceived to be, and were, of higher quality (i.e. more likely to retrieve relevant content) [38].

**Q3** Searchers, when primed with query auto-completions to promote more critical thinking, did not engage longer with the search task (as hypothesized by Yamamoto and Yamamoto [81]).

**Q4** More educated searchers, when primed with query auto-completions to promote critical thinking, issued more queries, checked the quality of sources more carefully, and selected higher quality articles that contained valid references [81].

**Q5** In terms of querying behaviour, it was hypothesized that a search system that presented results that were inconsistent with the searcher's beliefs would lead to lower search engagement and less querying. However, they found that searchers issued more queries and spent longer searching, trying to find information to confirm their beliefs [63]. Note that the search results were manipulated to return answers not consistent with the searchers beliefs, even if those answers were factually incorrect. The repeated querying and exposure to the alternative view, tended to sway searchers beliefs (suggesting *Exposure Effects*)[63] .

**Q6** When searchers were presented with suggested questions via a "*People also asked*" component, which provided answers inconsistent with searcher's prior beliefs, it led to search disengagement, fewer queries, less exploration of documents, and less time searching. So rather than mitigating *Confirmation Bias* as hypothesized by Pothirattanachaikul et al. [62], it actually led to *Information Avoidance*.

## 3.4 Cognitive Biases when Examining Lists

A phenomena that has been observed in many contexts, but in particular, during web search, is what is referred to by system sided researchers as **Position Bias** [13, 31]. This is where highly ranked

[3]It should be noted that participants in the study were deciding among candidates from another country running for an election that they had no vested interest in and could not vote in. Also, they could not search for other information about the candidates, only examine the fabricated/mock result list presented to them.

documents tend to attract more clicks than results ranked lower – and so there is an exponential decrease in the number of clicks as rank increases. When examining result lists other biases have also been noted, such as:

- **Attractiveness Bias** where searchers tend to prefer more attractive results over less attractive [82],
- **Domain Bias** where searchers judge results more favourably from known domains (reducing uncertainty) [29, 34], and,
- **Authority/Trust Bias** where searchers tend to trust the search engine rankings because they come from a perceived authority [57, 58].

Researchers have naturally considered whether these biases observed on the system side (which lead to evaluation issues and algorithmic biases where the *rich-get-richer* [12]), stem from cognitive biases on the user side [7]. Of interest, has been whether searchers follow Zipf's Law [84], the *Principle of Least Effort*, or not, and, if not, whether the observed click distribution is due to *Primacy* and *Recency Order Effects* [11, 80], *Authority/Trust Bias* [58], or *Satisficing* [35]. Typically, researchers investigating this phenomena, perform experiments where results items are swapped, or result lists are reversed, to tease out the different effects [11, 35, 58].

Pan et al. [58] claim that *Position Bias* arises because results are ranked in decreasing ordering of their probability of relevance [64], and over time search engines have learnt to rank the most relevant items first [31]. They suggest that taking the first result item returned is an example of a "*fast and frugal heuristic that exploits the regularity of the information environment*" [58]. So the trust people have in the search engine is proportional to the probability of result items being relevant, which is learnt from repeated interactions with the search engine. Thus, Pan et al. [58] assert that searcher's trust in the system is not unfounded or irrational. Instead they posit that by favouring the highly ranked items because they are more likely to be relevant will reduce the total amount of time and effort locating relevant information. Keane et al. [35] argue that searchers are satisficers: and will click on the first item or first relevant looking item that they encounter, rather than being maximizers who compare all results items before clicking on the best item. In line with the assertion by Pan et al. [58], Keane et al. [35] also suggest that the position bias is dependent on relevance — and when the relevance of the top results is reduced they observed that people are less likely to click them — suggesting that they do not always (or all) naively click the first result.

On the other hand, Wu et al. [80] contest this view, and suggest that if people abide by Zipf's Law, then they would seek to obtain the maximum benefit with the minimum effort. And so, when information seeking, Zipf-like distributions should result because of the trade-off between the benefit and effort, where people are less likely to seek information if it requires more effort to obtain it. Wu et al. [80] posit that the click distribution is not strictly Zipfian due to *Primacy* and *Recency Position Effects*. They observed that the first link receives the most clicks, but the tenth link on the page receives more that the 9th link, while the 11th link which is on the subsequent page receives more clicks than the 10th or 9th. Their statistical analysis of web search log data shows that the click distribution actually violates Zipf's Law which they attribute to *Primacy and Recency Effects*. We summarize these findings and hypotheses below:

**E1** *Position Bias* is proportional to the likelihood of results being relevant [35, 58]. If the relevance of top results decreases, searchers are less likely to click them [35].

**E2** It is hypothesized that searchers minimize the cost of their search by favouring high ranked items (and thus, searchers should follow Zipf's Law) [58].

**E3** However, if searchers exhibit *Primacy* and *Recency Order Effects* then it will lead to deviations from Zipf's Law [80].

**E4** When searching for answers to factoid questions like "*Who invented java?*" people tend to *satisfice* and take the first result with a credible answer, rather than trying to find the best possible result (i.e. *maximise*)[4]. [35].

### 3.5 Cognitive Biases when Judging Relevance

Commissioning relevance assessments to train and evaluate Information Retrieval systems is another area where researchers have been concerned that (cognitive) biases may have an impact on the quality of judgements (which may then have downstream effects on training and evaluating ranking models). For most studies, paid assessors are used to perform the annotations given a set of guidelines – and in most studies, these assessors are paid per label, and they are only paid, if their work is of sufficient quality – this creates an incentive structure which might exacerbate or interact with cognitive biases. These annotation studies have found that:

**A1** Assessors tend to rate more well known sites like Wikipedia as more relevant [29, 34] (referred to as *Domain Bias* [29] but can be attributed *Ambiguity Bias* as unknown sites have more uncertainty associated with them).

**A2** Assessors rate more detailed result summaries as more relevant than results that are missing features such as title, domain, etc. (*Ambiguity Bias*) [17]

**A3** Priming assessors by showing them examples of low relevance items, tended to result in assessors rating subsequent items more highly, than if they were shown highly or moderately relevant items first (*Priming Effects*) [67, 69].

**A4** After assigning a relevance label to the result summary, assessor tended to anchor on their initial judgement when assessing the document (*Anchoring Bias*) – and this resulted in lower accuracy compared to gold standard labels [17].

**A5** When two non-relevant items were shown with a relevant item, but one non-relevant item was clearly inferior to the other (i.e. the decoy), then assessors tended to rate the dominate non-relevant document as more relevant (*Decoy Effect*) [17].

**A6** When information on previous assessor decisions was provided (e.g. % of assessors who said the item was relevant) along with the result summary, the assessors tended to rate the document in a similar fashion, even if the % data was fabricated (*Anchoring Bias* and *Bandwagon Effects*) [17, 27].

[4]However, it is not really clear what the best possible result would be in this context since to complete the task participants only had to find a result with the answer to for instance, "*Who invented Java?*", and not say a more information result that contained the entire history of Java, which may be more informative.

**A7** When actual/true information regarding previous assessors decisions was provided (e.g. the true % of assessors who said the item was relevant), a stronger effect was observed such that: as the true % increased, then the probability of a new assessor labelling the item as relevant also increased proportionally (*Bandwagon Effects*) [27]. This suggests that when the information is believed to be genuine or is more credible it is more likely to influence assessor decisions.

**A8** When actual/true information was provided, the accuracy of judgements increased and assessors were more confident in their judgements [27].

**A9** When training ranking algorithms using the labels from assessors who used interfaces designed to induce cognitive biases, it resulted in lower performance compared to using labels, where assessors used a neutral interface[5] [17].

## 3.6 Cognitive Biases on Search Satisfaction

It has been shown that commonly used metrics in Information Retrieval, which are based on Expected Utility, tend to correlate poorly with search satisfaction [2]. Cognitive biases may be responsible for this deviation, however, there has been little work examining how cognitive biases impact upon satisfaction [39, 49, 56].

Kelly et al. [39] investigated whether priming users by providing feedback on their performance with the system influenced their satisfaction ratings. When told they performed very well (i.e. better than what was measured), ratings increased, while when told they performed poorly, ratings decreased, suggesting that participants were susceptible to the priming. However, they also observed that ratings were higher when no feedback was given compared to when their actual performance was given as feedback. They posit that *Method Bias* stemming from the experimental setting and the experimental design (i.e. participants are performing a task in an artificial environment) [60] may explain why participants give higher ratings on average (resulting in an *Inflation Bias*). This inflation bias could be problematic when comparing ratings between systems, if the relative differences are not proportional to the ratings. However, when novel search interfaces, with clear interventions, are compared to baseline interfaces, participants may fall prey to *Novelty Bias* [15], leading to a disparity in the inflationary effects.

Oulasvirta et al. [56] investigated the *Paradox of Choice* [68] in the context of web search. In their study, they presented participants with search result lists containing either six result items or twenty four result items, from which participants had to select one result item. While search result lists of other sizes were not examined, they hypothesized that an inverted U-Shaped relationship between the number of results and satisfaction exists, such that too few choices or too many choices would lead to greater dissatisfaction. They found that participants reported greater satisfaction, greater confidence and greater consideration of items when result pages contained six items, suggesting that less is more. However, the experimental setting put participants into a *maximiser* vs *satisficer* scenario, forcing them to select one, and only one result item (similar to the assumption made by Keane et al. [35], see §3.4). In a naturalistic setting, where users may wish to click on and inspect a number of items, this may not necessarily be the case, and it may even increase the cost of browsing [6, 37].

More recently, Liu and Han [49] investigated whether *Reference Dependence Effects* influenced participants satisfaction. Reference Dependence effects manifest when people view gains and losses with respect to a reference point, rather than the absolute gains or losses (as assumed under Expected Utility theory). They examined several datasets where participants had performed search tasks and provided satisfaction ratings. They found evidence to suggest that participant's satisfaction was significantly influenced by their peak gain or peak loss, and their final end gain or loss (*Peak-End Rule*).

**S1** Participants tend to rate experimental systems more highly, especially if the system is novel (*Inflation and Novelty Biases*) [39].

**S2** When participants have to choose only one item, they were less satisfied when many results were shown, than when fewer results were shown (*Less is More Effect*) [56].

**S3** Peak and end events act as reference points which can influence participant's search satisfaction ratings (*Peak-End Rule)* [49].

# 4 DISCUSSION AND CRITICAL REFLECTION

In the previous sections we have highlighted some of the major findings observed across different domains and different parts of the search process – it is clear that cognitive biases influence people's information seeking and retrieval behaviours and the subsequent decisions that they make to some degree or another – and these cognitive biases can be exacerbated when interacting with algorithmically biased search engines that serve up biased content. However, it should be pointed out that the above findings are tendencies, which may lead to an increase in errors beyond what is expected according to rational decision making theories (but not always and not for all searchers). Most of the studies in Table 1 have focused on examining the negative impacts of cognitive biases on search. However, cognitive biases can also have a positive impact (**A7**, **Q4**), which can potentially and presumably result in more successful outcomes in the long run (**E1**, **E2**) — however, further work is needed to investigate the positive impacts. Below, we discuss a number of open questions, issues and challenges in investigating cognitive biases in information seeking and retrieval.

**Do cognitive biases compound with each other or conflict?** The cognitive biases which seem to have the biggest impact appear to be *Anchoring Bias*, *Confirmation Bias*, *Ambiguity Effects* and *Exposure Effects* which have been observed in different contexts influencing searchers in similar ways. It also appears that certain biases are hard to distinguish because they are often observed together, and potentially lead to compounding effects. For example, *Availability Bias* and *Exposure Effects* tend to be linked, because the availability created by presenting documents with a certain point of view in the result list, coupled with a searcher engaging with a number of those results, can lead to a shift or sway in the searcher's opinions and beliefs (**HM5**, **SP5**, **SP6**). While *Confirmation Bias* and *Information Avoidance Bias* tend to combine together as searchers avoid conflicting opinions, but will seek out opinions that are consistent with their beliefs (**HM8**, **SP1**). *Anchoring Bias* and *Primacy Bias* also tend to co-occur, or at least are conflated with each other, as the first item presented in a result list is often

[5]Performance was in terms of nDCG calculated using "gold" labels from NIST assessors.

considered the most relevant, and may serve as an anchor that influences how searchers answer questions (**HM3**, **E1**). We also observed that there can be a tension between different cognitive biases. For example, despite participants anchoring towards a particular belief and performing repeated searches to try to confirm their beliefs, they succumbed to exposure effects as the search engine deliberately returned results inconsistent with their beliefs (**H6**). Participants who made an initial decision on who to vote for (anchor) were also swayed by *Exposure and Primacy Effects*. However, when assessors provided an initial judgement, they stuck to their initial anchor despite other manipulations (**A4**). As such, a number of open questions arise when considering cognitive biases in ISR settings: Which of these cognitive biases compound together and lead to better or worse outcomes? Which biases compete against each other? Do they wash each other out? Or, do certain cognitive biases dominate over others?

**Is it easier to manipulate or mitigate cognitive biases?** Some answers can be found by examining the findings from studies that have tried to mitigate different cognitive biases. These works suggest that some biases appear to be more deeply ingrained, or that the intervention was not appropriate to mitigate the bias. For example, while health search interfaces could help minimize *Order Effects*, searchers *a priori* beliefs were harder to overcome (e.g. Anchoring Bias), whereas trying to mitigate *Confirmation Bias* actually led to *Information Avoidance Bias* (**Q6**). On the other hand, studies in which researchers have deliberately created interfaces and scenarios to identify or detect different cognitive biases have been largely successful (**HM3**, **HM5**, **SP6**, **A3-6**, **E3**, **S2**). These findings suggest that it is harder to mitigate cognitive biases and their negative effects, than it is to manipulate cognitive biases and exacerbate their negative effects. However, greater manipulations also require greater experimentation control which limits their ecological validity. For example, in [19], they claim that undecided voters may be swayed by 20% or more towards a particular candidate (**SP6**). However, participants were only presented with one list of results where the ordering had been manipulated (to favour one or another candidate). Participants could not issue their own queries – so even if they wanted to they were unable to take other actions that may have mitigated the biases imposed by the lists presented to them. On the other hand, providing initial results which contain factually correct answers help to improve the accuracy of searchers when answering medical questions (**HM3**, **HM5**). Essentially, it is very much an open question how much influence search engine manipulations can actually effect searcher opinions, beliefs and decisions in naturalistic and ecologically valid settings. But worryingly, it appears that leading searchers to incorrect/poorer conclusions by manipulating their cognitive biases, is easier than mitigating cognitive biases to help lead searchers to the correct/better conclusions.

**Are searchers more susceptible to cognitive biases if they have no skin in the game?** It may be that the effects observed in the controlled and artificially constructed experimental scenarios may be because the participants are not intrinsically motivated to find out whether, for example, "cinnamon helps diabetes" or not. However, if a person has been recently diagnosed with Type II Diabetes, then they have a clear motivation for learning about their condition, the possible treatments and seeking advice on dealing with the condition. In this case, would they really satisfice and take one and only one answer provided by the search engine (**E3**, **S2**)? More likely, they will undertake many search sessions, issue many queries, examine many different articles, as well as discuss issues with their friends, family and medical providers. Of course, they may avoid certain results when searching ( *Information Avoidance Bias*) or only select results that confirm their prior beliefs (*Confirmation Bias*), but it is largely unknown what the actual influence would be when the searcher's health is at stake. Another possibility may be that some/most participants in these experimental studies may be looking to do the experiment as quickly as possible, so they can claim the payment and leave. In this case the motivations are not necessarily aligned with the experimental context, and this could lead to observing more cognitive biases, or more pronounced effects from such cognitive biases. But perhaps reassuringly, the more effort the participants put into their search task, the more accurate their answers (**HM7**) suggesting that if they do try harder they are more likely to reach a better outcome/correct answer. Nonetheless, these points emphasize that care needs to be taken when designing and conducting such experiments, so that they aren't subject to *Method Bias* [60] or the *Observer-Expectancy Effects* [65].

**On that note, are we, as ISR researchers, also cognitively biased when performing studies on cognitive bias in ISR?** It is important to take a step back and ask this meta-question, and consider whether we as researchers are falling foul to the *Observer-Expectancy Effect* [65]. This is where experimenters can suffer a form of *Confirmation Bias*, incorrectly interpreting results to see patterns and trends that conform to their hypotheses, rather than considering alternative hypotheses. In the following discussion points, we critically examine some of the past findings and consider whether there may be alternative explanations for the results observed in previous studies as a way to highlight the difficulties in determining whether the influence is due to cognitive biases or due to some other cause. Taking, for example, some of the findings from studies on health searchers, where it was suggested that people tend to issue positively framed queries (**HM1**), such as "*Does cinnamon help diabetes?*", and then select confirmatory results (**H2**), which in turn increases their likelihood of giving incorrect answers (**H3**,**H5**). But, are there other explanations or other factors that might account for these observations?

**Are positively framed queries an example of projection bias?** Over time searchers have interacted with search engines which has resulted in learning what kinds of queries tend to work, vs. what kinds of queries don't work (through *Operant Conditioning* [71]). Predominately, people are conditioned by the search engines to favour keyword based queries such that searchers find documents which contain those keywords, where as more complex queries involving "not" or "NOT" operators are often misunderstood and frequently used incorrectly. So perhaps the searcher by asking queries is not so much projecting their belief that "*cinnamon does help diabetes*" but actually, thinks that this query will retrieve relevant information to address their underlying information need. Maybe, they are suffering from a different cognitive bias? Or, maybe they lack the domain knowledge to formulate a better query? For example, it is probably cognitively less taxing to ask the search engine: "*Does caffeine help asthma?*" and read casually written health articles, compared to formulating a more expressive and balanced query like "*Randomized controlled clinical trials on exercise induced bronchoconstriction asthma participants involving caffeine as a treatment*" and read complex scientific articles (**HM9**, **SP5**).

**Are searchers exhibiting confirmation bias because they select documents claiming the treatments are helpful?** What makes answering this question particularly difficult in a naturalistic setting (such as via web search logs) is that there is a high proportion of documents in the index that contain positive answers. In fact, up to 80% of top ranked documents disproportionately favour such results due to *Content Bias* [78]. This system-sided bias may contribute to why searchers selected documents with positive answers, as that is what was mainly shown, rather than the searchers wanting to confirm their beliefs. However, in more controlled experiments, similar findings have been observed, suggesting that searchers are often being selective in the items that they are examining, favouring ones that are consistent with their beliefs and, thus, exhibiting *Confirmation Bias* (SP1) .

**Did searchers have a different interpretation of the search task?** As previously mentioned, researchers used commonly posed health questions, for which there was a clinical answer based on Cochrane systematic reviews (the gold standard). For instance, cinnamon is NOT an effective treatment for diabetes – and thus was deemed as not helpful. So, if a participant said it was helpful, their answer was considered incorrect by the researchers [46, 61]. This is because they assume that the participant's understanding of the search task, and the question that they had to answer, was the same as theirs. However, participants may conclude cinnamon is helpful because it can be used as substitute for sugar. Similarly, while Aloe Vera is NOT a cure for cancer, participants may have concluded it is helpful because it is effective in alleviating the pain caused by chemotherapy burns. So, how much of the observed inaccuracies are due to cognitive biases vs. misinterpretations of the search task?

**Do misaligned incentive structures lead assessors to employ simple rules, or are they cognitively biased?** When commissioning annotations, the annotation platform would like high quality, unbiased, accurate labels for the minimum cost. However, the assessors motivations are likely to be very different – earn as much money as possible for the least amount of effort. Clearly, the assessor's motivations differ from the platform's objectives. Furthermore, it maybe that the annotation platform and the incentive structures that it provides encourages assessors to take (mental) short cuts. Let's say that an assessor decides to universally employ a simple rule, for example, to always label a item based on what the majority thought given the platform provides this information. Now, let's further assume that this rule leads to correctly labelling the majority of labels most of the time (as per **A7** and **A8**). Clearly, employing this rule would then be beneficial to the assessor – as they would be able to quickly label items reasonably accurately, and earn a higher hourly wage (as many annotation platforms pay per label, assuming the labelling is sufficiently accurate). On the other hand, if this rule led to low accuracy, then they run the risk of not meeting the quality threshold, and may not get paid. Essentially, they would only apply this rule if it pays off. So a savvy assessor is likely to game the system by creating simple rules to perform the annotations quickly (e.g. rate results from well known sites higher, if popular, rate higher, etc.), but do so consciously, as a way to deliberately maximise their income (via system 2 thinking). On the other hand, they may be unconsciously learning to quickly perform annotations (via system 1 thinking). However, both would appear to give the same empirical observations, but arise due to different ways of thinking. So, are assessors jumping on the bandwagon, or are they gaming the system?

**Are searchers satisficers, maximisers or optimisers?** In the context of assessing result lists, it has been argued that searchers are *satisficers* rather than *maximisers* (**HM9**, **E3**) – that is they tend to select the first acceptable result, rather than compare all the results and select the very best one. It was argued that this behaviour is rational because searchers are subscribing the *Principle of Least Effort* – and this leads to observing a Zipfian-like distribution of clicks because searchers are less likely to seek information if more effort is required. While Zipf's law is not a perfect fit attributed to *Primacy* and *Recency Effects* (**E3**), it does suggest that people are trying to optimise their ISR. And, rather than trying to be a *maximiser* by selecting the best possible result from a result list, it may be that they are trying to be an *optimiser* where they seek to maximise their rate of gain — which is one of the key assumptions behind Information Foraging Theory [59] and other related models of search based on Expected Utility [3, 4, 21]. However, there are still many open questions: what objective function do searchers employ? How does it change depending on the context? To what extent do cognitive biases influence their objective function? And, do certain cognitive biases lead to more efficient and/or effective ISR, or not?

While it is impossible to discuss all the different issues that have emerged during our review of the literature on cognitive biases in search, we hope that these discussion points help the field to: critically reflect upon the work being performed, highlight the challenges in investigating this phenomena, and provide a strong starting point for researchers wanting to explore cognitive biases in future work.

## 5 SUMMARY AND CONCLUDING REMARKS

Cognitive Biases can play a major role in shaping and influencing ISR behaviours and outcomes. This perspectives paper has outlined the major works conducted in our field, cataloguing and describing the different findings and assertions from over thirty studies. From this review, it is clear that we, as a field, have only just scratched the surface in understanding the complexities of cognitive biases in ISR. And, that there are many challenges and open questions that need to be addressed and answered going forward. We hope that this work serves as inspiration for investigating the many complexities of this phenomena exploring, for example, the impact of compounding and conflicting cognitive biases, the negative *and* positive effects of cognitive biases, and how cognitive biases can be mitigated (and manipulated) to improve ISR behaviors and outcomes.

**Acknowledgements**: I would like to thank to Dr. David Maxwell, Dr. Guido Zuccon and Molly McGregor for their insightful discussions and helpful feedback in preparing this work. I would also like to thank the reviewers for their suggestions and feedback.

## REFERENCES


[1] Mustafa Abualsaud and Mark D Smucker. 2019. Exposure and Order Effects of Misinformation on Health Search Decisions. In *ROME 2019 Workshop on Reducing Online Misinformation Exposure*. ACM.

[2] Azzah Al-Maskari and Mark Sanderson. 2010. A review of factors influencing user satisfaction in information retrieval. *Journal of the American Society for Information Science and Technology* 61, 5 (2010), 859–868. https://doi.org/10.1002/asi.21300

[3] Leif Azzopardi. 2011. The Economics in Interactive Information Retrieval. , 15–24 pages. http://10.0.4.121/2009916.2009923

[4] Leif Azzopardi. 2014. Modelling interaction with economic models of search. In *SIGIR 2014 - Proceedings of the 37th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 3–12. https://doi.org/10.1145/2600428.2602298

[5] Leif Azzopardi and Vishwa Vinay. 2008. Retrievability: An evaluation measure for higher order information access tasks. In *International Conference on Information and Knowledge Management, Proceedings*. 561–570. https://doi.org/10.1145/1458082.1458157

[6] Leif Azzopardi and Guido Zuccon. 2016. Two scrolls or one click: A cost model for browsing search results. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 9626 (2016), 696–702. https://doi.org/10.1007/978-3-319-30671-1{_}55

[7] Ricardo Baeza-Yates. 2018. Bias on the web. *Commun. ACM* 61, 6 (2018), 54–61. https://doi.org/10.1145/3209581

[8] Nick Bansback, Linda C. Li, Larry Lynd, and Stirling Bryan. 2014. Exploiting order effects to improve the quality of decisions. *Patient Education and Counseling* 96, 2 (2014), 197–203. https://doi.org/10.1016/j.pec.2014.05.021

[9] Sara Behimehr and Hamid R. Jamali. 2020. Cognitive biases and their effects on information behaviour of graduate students in their research projects. *Journal of Information Science Theory and Practice* 8, 2 (2020), 18–31. https://doi.org/10.1633/JISTaP.2020.8.2.2

[10] Buster Benson. 2016. Cognitive bias cheat sheet. https://medium.com/better-humans/cognitive-bias-cheat-sheet-55a472476b18

[11] Keith Burghardt, Tad Hogg, and Kristina Lerman. 2018. Quantifying the impact of cognitive biases in question-answering systems. In *12th International AAAI Conference on Web and Social Media, ICWSM 2018*. 568–571.

[12] Junghoo Cho and Sourashis Roy. 2004. Impact of Search Engines on Page Popularity. In *Proceedings of the 13th International Conference on World Wide Web (WWW '04)*. ACM, 20–29. http://doi.acm.org/10.1145/988672.988676

[13] Nick Craswell, Onno Zoeter, Michael Taylor, and Bill Ramsey. 2008. An experimental comparison of click position-bias models. In *Proceedings of the 2008 International Conference on Web Search and Data Mining*. ACM, Palo Alto, California, USA, 87–94. https://doi.org/10.1145/1341531.1341545

[14] Robert G Crowder. 2014. *Principles of learning and memory: Classic edition* (classic ed ed.). Psychology Press.

[15] Nicola Dell, Vidya Vaidyanathan, Indrani Medhi, Edward Cutrell, and William Thies. 2012. "Yours is better!": participant response bias in HCI. In *Proceedings of tje Conference on Human Factors in Computing Systems*. ACM, 1321–1330. https://doi.org/10.1145/2207676.2208589

[16] David Dunning. 2011. The dunning-kruger effect. On being ignorant of one's own ignorance. *Advances in Experimental Social Psychology* 44 (2011), 247–296. https://doi.org/10.1016/B978-0-12-385522-0.00005-6

[17] Carsten Eickhoff. 2018. Cognitive biases in crowdsourcing. In *WSDM 2018 - Proceedings of the 11th ACM International Conference on Web Search and Data Mining*. 162–170. https://doi.org/10.1145/3159652.3159654

[18] David Elsweiler, Christoph Trattner, and Morgan Harvey. 2017. Exploiting food choice biases for healthier recipe recommendation. In *SIGIR 2017 - Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 575–584. https://doi.org/10.1145/3077136.3080826

[19] Robert Epstein and Ronald E. Robertson. 2015. The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. *Proceedings of the National Academy of Sciences (PNAS)* 112, 33 (2015), E4512–E4521. https://doi.org/10.1073/pnas.1419828112

[20] Deborah Frisch and Jonathan Baron. 1988. Ambiguity and rationality. *Journal of Behavioral Decision Making* 1, 3 (1988), 149–157. https://doi.org/10.1002/bdm.3960010303

[21] Norbert Fuhr. 2008. A probability ranking principle for interactive information retrieval. *Information Retrieval* 11, 3 (2008), 251–265.

[22] Ujwal Gadiraju and Stefan Dietze. 2017. Improving learning through achievement priming in crowdsourced information finding microtasks. In *LAK '17: Proceedings of the Seventh International Learning Analytics & Knowledge Conference*. 105–114. https://doi.org/10.1145/3027385.3027402

[23] Amira Ghenai, Mark D. Smucker, and Charles L.A. Clarke. 2020. A think-aloud study to understand factors affecting online health search. In *CHIIR 2020 - Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*. 273–282. https://doi.org/10.1145/3343413.3377961

[24] Gerd Gigerenzer and Henry Brighton. 2009. Homo Heuristicus: Why Biased Minds Make Better Inferences. *Topics in Cognitive Science* 1, 1 (1 2009), 107–143. https://doi.org/10.1111/j.1756-8765.2008.01006.x

[25] Gerd Gigerenzer and Richard Selten. 1999. *Bounded Rationality: The Adaptive Toolbox*. MIT Press.

[26] Jeannine Cyr Gluck. 2020. How Searches Fail: Cognitive Bias in Literature Searching. *Journal of Hospital Librarianship* 20, 1 (2020), 27–37. https://doi.org/10.1080/15323269.2020.1702839

[27] Christopher G. Harris. 2019. Detecting cognitive bias in a relevance assessment task using an eye tracker. In *ETRA '19: Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*. 1–5. https://doi.org/10.1145/3314111.3319824

[28] Joel Huber, John W Payne, and Christopher Puto. 1982. Similarity Hypothesis Linked references are available on JSTOR for this article : Adding Asymmetrically Dominated Alternatives : Violations of Regularity and the Similarity Hypothesis. *Journal of Consumer Research* 9, 1 (1982), 90–98.

[29] Samuel Ieong, Nina Mishra, Eldar Sadikov, and Li Zhang. 2012. Domain Bias in Web. In *WSDM 2012 - Proceedings of the 5th ACM International Conference on Web Search and Data Mining*. ACM, 413–422.

[30] Peter Ingwersen and Kalervo Järvelin. 2005. *The Turn: Integration of Information Seeking and Retrieval in Context (The Information Retrieval Series)*. Springer-Verlag New York, Inc.

[31] Thorsten Joachims. 2002. Optimizing search engines using clickthrough data. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 133–142. https://doi.org/10.1145/775047.775067

[32] Daniel Kahneman. 2017. Thinking, fast and slow.

[33] Markus Kattenbeck and David Elsweiler. 2019. Understanding credibility judgements for web search snippets. *Aslib Journal of Information Management* 71, 3 (2019), 368–391. https://doi.org/10.1108/AJIM-07-2018-0181

[34] Gabriella Kazai, Nick Craswell, Emine Yilmaz, and S. M.M. Tahaghoghi. 2012. An analysis of systematic judging errors in information retrieval. In *Proceedings of the 21st ACM international conference on Information and knowledge management*. 105–114. https://doi.org/10.1145/2396761.2396779

[35] Mark T. Keane, Maeve O'Brien, and Barry Smyth. 2008. Are people biased in their use of search engines? *Commun. ACM* 51, 2 (2008), 49–52. https://doi.org/10.1145/1314215.1314224

[36] Diane Kelly. 2009. Methods for Evaluating Interactive Information Retrieval Systems with Users. *Foundations and Trends in Information Retrieval* 3 (2009), 1–224.

[37] Diane Kelly, Jaime Arguello, Ashlee Edwards, and Wan Ching Wu. 2015. Development and evaluation of search tasks for IIR experiments using a cognitive complexity framework. In *ICTIR 2015 - Proceedings of the 2015 ACM SIGIR International Conference on the Theory of Information Retrieval*. ACM, 101–110. https://doi.org/10.1145/2808194.2809465

[38] Diane Kelly, Amber Cushing, Maureen Dostert, Xi Niu, and Karl Gyllstrom. 2010. Effects of popularity and quality on the usage of query suggestions during information search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Vol. 1. 45–54. https://doi.org/10.1145/1753326.1753334

[39] Diane Kelly, Chirag Shah, Cassidy R. Sugimoto, Earl W. Bailey, Rachael A. Clemens, Ann K. Irvine, Nicholas A. Johnson, Weimao Ke, Sanghee Oh, Anezka Poljakova, Marcos A. Rodriguez, Megan G. Van Noord, and Yan Zhang. 2008. Effects of performance feedback on users' evaluations of an interactive IR system. *IIiX'08: Proceedings of the 2nd International Symposium on Information Interaction in Context* (2008), 75–82. https://doi.org/10.1145/1414694.1414712

[40] Diane Kelly and Cassidy Sugimoto. 2013. A Systematic Review of Interactive Information Retrieval Evaluation Studies, 1967-2006. *Journal of the American Society for Information Science and Tech.* 64, 4 (2013), 745–770.

[41] Alla Keselman, Allen C. Browne, and David R. Kaufman. 2008. Consumer Health Information Seeking as Hypothesis Testing. *Journal of the American Medical Informatics Association* 15, 4 (2008), 484–495. https://doi.org/10.1197/jamia.M2449

[42] Joshua Klayman and Young-won Ha. 1987. Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review* 94, 2 (1987), 211–228. https://doi.org/10.1037//0033-295x.94.2.211

[43] Silvia Knobloch-Westerwick, Benjamin K. Johnson, and Axel Westerwick. 2015. Confirmation bias in online searches: Impacts of selective exposure before an election on political attitude strength and shifts. *Journal of Computer-Mediated Communication* 20, 2 (2015), 171–187. https://doi.org/10.1111/jcc4.12105

[44] Juhi Kulshrestha, Motahhare Eslami, Johnnatan Messias, Muhammad Bilal Zafar, Saptarshi Ghosh, Krishna P. Gummadi, and Karrie Karahalios. 2017. Quantifying search bias: Investigating sources of bias for political searches in social media. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*. 417–432. https://doi.org/10.1145/2998181.2998321

[45] Sanna Kumpulainen and Hugo Huurdeman. 2015. Shaken, not steered: The value of shaking up the search process. *CEUR Workshop Proceedings* 1338 (2015), 1–4.

[46] Annie Y.S. Lau and Enrico W. Coiera. 2007. Do People Experience Cognitive Biases while Searching for Information? *Journal of the American Medical Informatics Association* 14, 5 (2007), 599–608. https://doi.org/10.1197/jamia.M2411

[47] Annie Y.S. Lau and Enrico W. Coiera. 2009. Can Cognitive Biases during Consumer Health Information Searches Be Reduced to Improve Decision Making?

*Journal of the American Medical Informatics Association* 16, 1 (2009), 54–65. https://doi.org/10.1197/jamia.M2557

[48] Richard L. Lewis, Andrew Howes, and Satinder Singh. 2014. Computational rationality: Linking mechanism and behavior through bounded utility maximization. *Topics in Cognitive Science* 6, 2 (2014), 279–311. https://doi.org/10.1111/tops.12086

[49] Jiqun Liu and Fangyuan Han. 2020. Investigating Reference Dependence Effects on User Search Interaction and Satisfaction: A Behavioral Economics Perspective. In *SIGIR 2020 - Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1141–1150. https://doi.org/10.1145/3397271.3401085

[50] Jamie Murphy, Charles Hofacker, and Richard Mizerski. 2006. Primacy and Recency Effects on Clicking Behavior. *Journal of Computer-Mediated Communication* 11, 2 (2006), 522–535. https://doi.org/10.1111/j.1083-6101.2006.00025.x

[51] Raymond S. Nickerson. 1998. Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology* 2, 2 (1998), 175–220. https://doi.org/10.1037/1089-2680.2.2.175

[52] Safiya Umoja Noble. 2018. *Algorithms of oppression: How search engines reinforce racism*. NYU Press.

[53] Alamir Novin and Eric Meyers. 2017. Making Sense of Conflicting Science Information: Exploring bias in the search engine result page. In *Proceedings of the 2017 conference on conference human information interaction and retrieval*. 175–184. https://doi.org/10.1145/3020165.3020185

[54] Alamir Novin and Eric M Meyers. 2017. Four Biases in Interface Design Interactions BT - Design, User Experience, and Usability: Theory, Methodology, and Management, Aaron Marcus and Wentao Wang (Eds.). Springer International Publishing, Cham, 163–173.

[55] Jörg Oechssler, Andreas Roider, and Patrick W Schmitz. 2008. Cooling-Off in Negotiations - Does It Work? *SONDERFORSCHUNGSBEREICH 504* May 2008 (2008).

[56] Antti Oulasvirta, Janne P Hukkinen, and Barry Schwartz. 2009. When more is less: the paradox of choice in search engine use. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*. ACM, Boston, MA, USA, 516–523. https://doi.org/10.1145/1571941.1572030

[57] Maeve O'Brien and Mark T. Keane. 2006. Modeling result-list searching in the world wide web: The role of relevance topologies and trust bias. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society*. 1881–1886. https://api.semanticscholar.org/CorpusID:18100844

[58] Bing Pan, Helene Hembrooke, Thorsten Joachims, Lori Lorigo, Geri Gay, and Laura Granka. 2007. In Google we trust: Users' decisions on rank, position, and relevance. *Journal of Computer-Mediated Communication* 12, 3 (2007), 801–823. https://doi.org/10.1111/j.1083-6101.2007.00351.x

[59] Peter Pirolli and Stuart Card. 1999. Information foraging. *Psychological Review* 106 (1999), 643–675.

[60] Philip M Podsakoff, Scott B MacKenzie, Jeong-Yeon Lee, and Nathan P Podsakoff. 2003. Common method biases in behavioral research: A critical review of the literature and recommended remedies. , 879–903 pages. https://doi.org/10.1037/0021-9010.88.5.879

[61] Frances A. Pogacar, Amira Ghenai, Mark D. Smucker, and Charles L.A. Clarke. 2017. The positive and negative influence of search results on people's decisions about the efficacy of medical treatments. In *ICTIR 2017 - Proceedings of the 2017 ACM SIGIR International Conference on the Theory of Information Retrieval*. Association for Computing Machinery, Inc, 209–216. https://doi.org/10.1145/3121050.3121074

[62] Suppanut Pothirattanachaikul, Takehiro Yamamoto, Yusuke Yamamoto, and Masatoshi Yoshikawa. 2020. Analyzing the effects of "people also ask" on search behaviors and beliefs. In *Proceedings of the 31st ACM Conference on Hypertext and Social Media, HT 2020*. 101–110. https://doi.org/10.1145/3372923.3404786

[63] Suppanut Pothirattanachaikul, Yusuke Yamamoto, Takehiro Yamamoto, and Masatoshi Yoshikawa. 2019. Analyzing the effects of document's opinion and credibility on search behaviors and belief dynamics. In *International Conference on Information and Knowledge Management, Proceedings*. 1653–1662. https://doi.org/10.1145/3357384.3357886

[64] Stephen E Robertson. 1977. The probability ranking principle in IR. *Journal of documentation* 33, 4 (1977), 294–304.

[65] Robert Rosenthal. 1976. Experimenter effects in behavioral research. (1976).

[66] Luca Russo and Selena Russo. 2020. Search engines, cognitive biases and the man–computer interaction: a theoretical framework for empirical researches about cognitive biases in online search on health-related topics. *Medicine, Health Care and Philosophy* 23, 2 (2020), 237–246. https://doi.org/10.1007/s11019-020-09940-9

[67] Falk Scholer, Diane Kelly, Wan Ching Wu, Hanseul S. Lee, and William Webber. 2013. The effect of threshold priming and need for cognition on relevance calibration and assessment. *SIGIR 2013 - Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval* (2013), 623–632. https://doi.org/10.1145/2484028.2484090

[68] Barry Schwartz, Andrew Ward, Sonja Lyubomirsky, John Monterosso, Katherine White, and Darrin R. Lehman. 2002. Maximizing versus satisficing: Happiness is a matter of choice. *Journal of Personality and Social Psychology* 83, 5 (2002), 1178–1197. https://doi.org/10.1037/0022-3514.83.5.1178

[69] Milad Shokouhi, Ryen W. White, and Emine Yilmaz. 2015. Anchoring and adjustment in relevance estimation. In *SIGIR 2015 - Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 963–966. https://doi.org/10.1145/2766462.2767841

[70] Herbert A Simon. 1955. A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics* 69, 1 (1955), 99–118. https://doi.org/10.2307/1884852

[71] Burrhus Frederic Skinner. 1965. *Science and human behavior*. Number 92904. Simon and Schuster.

[72] Betsy Sparrow, Jenny Liu, and Daniel M Wegner. 2011. Google Effects on Memory: Cognitive Consequences of Having Information at Our Fingertips. *Science* 333, 6043 (2011), 776–778. https://doi.org/10.1126/science.1207745

[73] Peter M Todd and Gerd Gigerenzer. 2000. Précis of" Simple heuristics that make us smart". *Behavioral and brain sciences* 23, 5 (2000), 727–741.

[74] Amos Tversky and Daniel Kahneman. 1974. Judgment under Uncertainty: Heuristics and Biases. *Science* 185, 4157 (1974), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

[75] Amos Tversky and Daniel Kahneman. 1992. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty* 5, 4 (1992), 297–323. https://doi.org/10.1007/bf00122574

[76] Ryen W White. 2013. Beliefs and Biases in Web Search. In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*. Dublin, Ireland, 3–12.

[77] Ryen W. White. 2014. Belief dynamics in web search. *Journal of the Association for Information Science and Technology* 65, 11 (2014), 2165–2178. https://doi.org/10.1002/asi.23128

[78] Ryen W. White and Ahmed Hassan. 2014. Content bias in online health search. *ACM Transactions on the Web* 8, 4 (2014). https://doi.org/10.1145/2663355

[79] Ryen W. White and Eric Horvitz. 2013. Captions and biases in diagnostic search. *ACM Transactions on the Web* 7, 4 (2013). https://doi.org/10.1145/2486040

[80] Mingda Wu, Shan Jiang, and Yan Zhang. 2012. Serial position effects of clicking behavior on result pages returned by search engines. In *CIKM '12: Proceedings of the 21st ACM international conference on Information and knowledge management*. ACM, 2411–2414. https://doi.org/10.1145/2396761.2398654

[81] Yusuke Yamamoto and Takehiro Yamamoto. 2018. Query priming for promoting critical thinking in web search. *CHIIR 2018 - Proceedings of the 2018 Conference on Human Information Interaction and Retrieval* 2018-March (2018), 12–21. https://doi.org/10.1145/3176349.3176377

[82] Yisong Yue, Rajan Patel, and Hein Roehrig. 2010. Beyond position bias: examining result attractiveness as a source of presentation bias in clickthrough data. In *Proceedings of the 19th international conference on World wide web*. ACM, Raleigh, North Carolina, USA, 1011–1018. https://doi.org/10.1145/1772690.1772793

[83] Yan Zhang. 2012. Consumer health information searching process in real life settings. In *Proceedings of the ASIST Annual Meeting*, Vol. 49. https://doi.org/10.1002/meet.14504901046

[84] George K Zipf. 1949. *Human Behavior and the Principle of Least-Effort*. Addison-Wesley.