International Journal of Information Management Data Insights 4 (2024) 100229

Contents lists available at ScienceDirect

International Journal of Information Management Data Insights

journal homepage: www.elsevier.com/locate/jjimei

Rational satisficing heuristics as determinants of online search behavior

Debora Di Caprio[a,*], Francisco J. Santos-Arteaga[b]

[a] *Department of Economics and Management, University of Trento, Trento, Italy*
[b] *Departamento de Análisis Económico y Economía Cuantitativa, Universidad Complutense de Madrid, Madrid, Spain*

ARTICLE INFO

Keywords:
Information retrieval
Online search
Satisficing
Click through rates
Uncertainty
Heuristics

ABSTRACT

We design a set of satisficing heuristic algorithms that mimic the online information retrieval behavior of rational decision makers (DMs) as reflected in their click through rates (CTRs). We illustrate how basic heuristic algorithms formalizing binary decision trees composed by 21 nodes and requiring DMs to observe one satisficing alternative lack the structural capacity to characterize the retrieval process. The algorithm requiring DMs to observe two satisficing alternatives –formalizing binary decision trees composed by 111 nodes – provides a sufficient approximation to their CTRs. Adding a third alternative – accounting for 351 nodes – delivers an almost identical set of CTRs to those displayed by DMs. The mimicking quality of the heuristic algorithms prevails as alternatives are added up to include the ten ranked within the first page of search results, incorporating 2,047 nodes to formalize the corresponding retrieval process. The set of algorithms bridges the gap between purely rational decision theory and heuristic behavior, illustrating how the CTRs observed can be generated by rational DMs performing a sequential search process while aiming to observe two or three satisficing alternatives. The decision-tree algorithmic structures presented are sufficiently malleable to introduce any potential modification to the beliefs and preferences of DMs and study its consequences in terms of CTRs.

1. Introduction

How much information do people set up to retrieve when performing a search online, and, more importantly, are searches purely random or follow a predetermined pattern? The initial answer to these questions was provided empirically and described users clicking on an average of two pages per search query (Jansen et al., 1998). More than twenty years later, the research across disciplines has focused on the reasons behind the clicking behavior of users but not on the actual complexity of their information retrieval processes (Miranda & Miah, 2023; Schmitt et al., 2018; Utku Özmen & Yucel, 2019; Zanganeh & Hariri, 2018).

However, this latter feature is extremely important for companies operating online since, besides the characteristics of the alternatives described in the snippets, the information retrieval capacity of users and subsequent complexity of the search process condition the expected success of their products. The intuition on which this idea is based follows from the search engine marketing literature (Rosário & Dias, 2023). Clearly, the ability of firms to locate their products among the initial results delivered by a search engine conditions their probability of success, a feature that justifies the prices paid for search engine optimization and marketing services.

In this regard, if decision makers (DMs) are willing to evaluate two alternatives out of those composing the first page of search results, the change in evaluation probability from shifting across ranking positions can be computed by comparing the corresponding click through rates (CTRs). This calculation is, at the same time, conditioned by the fact that DMs set out to observe two satisficing products. If DMs were to shift from two to three satisficing alternatives due to a modification in their information retrieval capacities or the features of the corresponding recommender system, the change in CTRs and subsequent clicking probabilities should be computable and far from trivial. This quality is complemented by the actual value of the different probabilities, which does not only affect the evaluation of a given alternative but those of the subsequent ones composing the decision tree.

The current paper addresses these questions and simulates sequential retrieval processes of different complexity calibrated to the actual CTRs displayed by the users so as to identify those that can replicate the actual behavior observed. The literature on decision systems follows a different methodology. It is typically based on the design of experimental scenarios used to compare the behavioral results obtained with the empirical ones observed among users (Speier-Pero, 2019). This analytical framework is also generally applied in the information systems

* Corresponding author at: Department of Economics and Management, University of Trento, Via Inama 5, 38122, Trento, Italy.
E-mail addresses: debora.dicaprio@unitn.it (D. Di Caprio), fransant@ucm.es (F.J. Santos-Arteaga).

https://doi.org/10.1016/j.jjimei.2024.100229

literature when focusing on e-commerce (Sun et al., 2020). A recent research trend utilizes behavioral data to extrapolate the main attributes of users and improve the training of deep-learning models designed to predict their CTRs (Li et al., 2020), which have been highlighted as an important categorization quality by computer scientists (Qin et al., 2020).

The current model proceeds in the opposite direction, designing a benchmark algorithm determined by the CTRs of online users that can then be empirically tested via experimental research. The evaluation structure defining the algorithm is sufficiently malleable to simulate the effect on the retrieval behavior derived from any modification in the preferences of DMs, providing a relevant framework of reference required in the information and expert systems literatures (Hong et al., 2021; Zhang et al., 2020).

The main objective of the paper is to define an algorithmic benchmark framework that can be used to simulate the behavior of different types of users – categorized by their information assimilation capacities – as they proceed through the alternatives ranked by a search engine. The algorithm encompasses the main features defining the characteristics of the alternatives and the information retrieval behavior of users, their preferences, and subsequent decisions. In this regard, the main intuition on which the design of the algorithm is based differs from the standard approach developed in the computer science literature (Wang et al., 2019). More precisely, the algorithm builds on a rational sequential decision model that generates information retrieval processes consistent with those observed in real-life settings (Advanced Web Ranking, 2021).

We define an intuitive decision theoretical model that relates the cutoff values determining the evaluation behavior of DMs to the distance in terms of semantic similarity between the descriptions observed in the snippets and the ideal ones considered by the user (Di Caprio et al., 2022a; Tavana et al., 2015a). The model assumes that, when performing a search, users do not have any information regarding the potential results obtained. If some predefined preferences exist, for instance, if users have a favorite website from where to acquire the information, then this predefined knowledge implies that they are not in an uncertain setting anymore but focus on a decision already made.

That is, we implicitly assume that if users have a favorite set of websites, they are part of the bookmarks defined in the corresponding browser. As a result, the model assigns a uniform convolution to the groups of variables composing the characteristics defining the alternatives. The formal model is kept sufficiently intuitive to highlight the complexity of the search process defined by DMs, who must consider combinations of the variables defining the characteristics, both observed and potential, together with the number of satisficing alternatives aimed to be observed.

The intuition provided by the decision theoretical model demonstrates how the information retrieval algorithm can incorporate the behavioral profiles of different types of DMs and describe their expected search patterns. In this regard, the resulting prescriptive environment can be used as a benchmark relative to which categorize DMs when training any of the machine learning techniques commonly used in the artificial intelligence literature.

The algorithm delivers a set of CTRs that replicate those observed in real-life settings. The benchmark framework provided by the algorithm allows to study the sequential modifications in the CTRs following from an increase in the information assimilation capacities of DMs. This type of scenario cannot be analyzed using trivial algorithms designed to mimic the CTRs but not to account for the interactions across observations and alternatives arising through the different search paths that may be defined by DMs as they retrieve information.

We illustrate the substantial differences existing between these trivial algorithms and the one introduced in the current paper, which is based on the different binary decision trees that can be defined by the alternatives composing the first page of results delivered by a search engine. In particular, we focus on the capacity of the algorithm to replicate the behavior of DMs as the complexity of their information retrieval processes is enhanced from a simple tree based on one alternative and consisting of 21 nodes to a complete one based on the whole set of alternatives and composed by 2047 nodes.

2. Literature review

The capacity of search engines to store large amounts of data has allowed researchers to analyze the online information retrieval behavior of users without facing the drawbacks derived from the observer bias phenomenon (Baclawski, 2018). The sequential evaluation of alternatives following the order of the ranking provided by search engines has been consistently recognized and validated through eye-tracking technology (Lorigo et al., 2008; Luo et al., 2011; Lewandowski & Kammerer, 2021; Wu & Yu, 2020).

Aside from small differences across time periods, the retrieval behavior of DMs is both consistent and biased towards the first two alternatives composing the ranking (Chitika, 2013; Dean, 2019). In addition, almost all clicks per search query are concentrated on the first page of results provided by the search engine (Advanced Web Ranking, 2021). This type of behavior relates to the substantial trust that DMs place on the rankings delivered by the engines (Epstein & Robertson, 2015; Guan & Cutrell, 2007; Silverstein et al., 1999).

The traditional approach to the information retrieval processes of DMs focuses on the behavior of individuals following the postulates of rational decision making (Di Caprio & Santos-Arteaga, 2021; Ren & Huang, 2018). However, the importance that the cognitive limits and assimilation capacities of DMs have on their retrieval behavior has been consistently highlighted in the literature (Basu, 2018; Joseph & Gaba, 2020; Victorelli et al., 2020; Yu et al., 2017). As a result, the satisficing approach to the behavior of boundedly rational DMs when retrieving information has gained relevance in experimental frameworks (Kabiawu et al., 2016; Lim, 2013; List & Alexander, 2017). This latter line of research has been complemented and validated through the analysis of empirical phenomena such as information overload (Walgrave & Dejaeghere, 2017) and choice regret (Schwartz, 2005).

Sequential search algorithms are therefore conditioned by the number of alternatives that users are willing to evaluate or the number of satisficing ones that they are aiming to find. The literature on information management has acknowledged the importance that the capacity of users to assimilate information has for the retrieval process. However, this feature has not been incorporated into the corresponding models.

For instance, Chen and Macredie (2010) reviewed the literature on human factors considered to be fundamental for the development of web applications, namely, gender, previous knowledge, and the cognitive qualities of users. In this regard, the development of recommender system is indeed based on the evidence available to extrapolate the preferences of users (Sharma et al., 2021) together with the relative valence of the topics describing the alternatives (Mishra et al., 2023). Similarly, the logical and semantic capacity of users has constituted one of the main challenges addressed when developing search engines (Shepherd, 2007). These constraints prevail nowadays and have been recently addressed through semantic query expansions within fuzzy ontology frameworks (Jain et al., 2021).

A similar intuition follows when considering an information supply perspective. For example, Kim et al. (2008) analyzed the importance that strategic positioning has for firm performance within an e-business context, a feature conditioning their relative position within the output results provided by a search engine. Sutanto et al. (2018) studied the impact that the retrieval behavior of knowledge workers had on work efficiency and concluded that retrieving knowledge from a self-made system did not foster efficiency due to the biased perceptions of workers.

All in all, the importance assigned to the cognitive capacities of DMs together with the empirical CTRs observed in real-life settings lead to the question of whether a rational sequential behavior is compatible with the cognitive limits imposed by the information retrieval processes

(Ham et al., 2019; O'Brien et al., 2020). We illustrate how both approaches, namely, a fully rational evaluation process and a heuristic mechanism limiting the number of alternatives considered by DMs, are compatible within a real-life information retrieval setting. The literature does not consider interactions between both approaches and tends to present them as excluding processes (Bossaerts & Murawski, 2015; Yang et al., 2015). We illustrate how rational sequential decision processes can be combined with heuristic mechanisms to identify and mimic the retrieval patterns observed in online search environment.

2.1. Contribution: how complex are online information retrieval processes?

We present a series of decision-tree algorithms of varying structural complexity based on increasingly demanding satisficing requirements assigned to DMs. The algorithms range from relatively simple frameworks composed by 21 nodes to complex structures accounting for the whole set of 2047 nodes that DMs may face as they retrieve information from the first page of results displayed by a search engine.

Fig. 1 describes the organic CTRs reported by Dean (2019) and those derived from international desktop searches in Advanced Web Ranking (2021). The similarity in the distribution of CTRs across alternatives through the different time periods is remarkable. We must highlight the fact that Dean (2019) does not consider clicks located in the second page of results due to their low CTRs, a quality verified by Advanced Web Ranking (2021), where the CTR of the alternative located in the eleventh position equals 0.95. The remaining alternatives composing the second page of results describe a descending tendency relative to this latter value.

The benchmark algorithm can be extended to incorporate additional alternatives, increasing the complexity of the subsequent decision trees considerably. The combinatorial requirements from adding one alternative to the analysis, namely, defining and simulating a tree composed by 4095 nodes, clearly outweigh the benefits derived from the additional explanatory capacity of the algorithm.

In addition to the satisficing heuristic, a unique constraint is imposed when designing the algorithms, namely, the trust placed by DMs in the ranking delivered by the search engine (European Commission, 2016; Pan et al., 2007).

We define a stochastic retrieval process where DMs face complete uncertainty regarding the alternatives provided by the engine after performing a search. A uniform distribution is used to formalize the lack of information over the characteristics of the alternatives, whose realizations determine the retrieval behavior derived from the search process. Thus, the nodes of the decision tree, which represent the alternatives displayed by the engine, are assigned uniform densities – though the results presented are independent of this assumption. Sets of

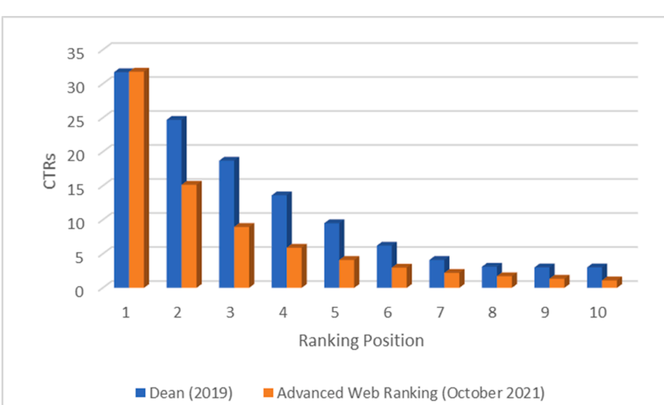

Fig. 1. CTRs across time periods: Dean (2019) versus Advanced Web Ranking (2021).

1,000,000 queries are simulated per algorithm to validate their capacity to mimic the retrieval behavior of DMs as reflected in their CTRs. In the current setting, CTRs represent the ratio of DMs who click on a specific alternative composing the ranking to the number of total queries performed.

Basic algorithms built on independent evaluations do not allow to incorporate the dynamic behavior of users observed in real-life environments, namely, going back to previous options after evaluating some of the alternatives, or just going forward and backward in the ranking to get an overall idea before deciding what to click. The benchmark algorithms introduced in the current paper account for this possibility, since they consider all the potential evaluation paths that may be defined by DMs as they proceed through the ranking delivered by the engine.

We summarize below the main contributions of the algorithmic framework defined in the current paper:

1. The algorithms are designed to describe a sequential information retrieval process that replicates the online search behavior of users when proceeding through a set of alternatives. That is, rational users are assumed to read the snippets, evaluate the information provided, and select the alternatives to click depending on the distance between the characteristics defining the alternatives – both observed and expected – and their preferred ones.
2. The algorithms incorporate each potential decision made together with the resulting paths that may be generated by DMs as they evaluate the alternatives composing the ranking. This feature allows to differentiate DMs according to the number of alternatives they aim to evaluate, which conditions the set of potential paths that may be followed through their retrieval processes.
3. The algorithms generate complete information retrieval processes based on the subjective preferences, beliefs, and assimilation capacities defining the behavior of DMs. The values of the CTRs observed empirically have been used to determine the probabilities assigned to the DMs clicking on the different alternatives.
4. In order to illustrate the complexity inherent to the formalization of the retrieval behavior of DMs, we introduce a simple algorithm that replicates the empirical CTRs of users but does not consider the sequential interactions that may arise among the evaluations and alternatives as DMs define the set of potential search paths when retrieving information.
5. The information assimilation capacities of DMs have been modified to illustrate the interactions arising through different sequential retrieval processes as well as the requirements necessary to generate the CTRs observed in real-life settings.

All in all, the subjective retrieval capacities of DMs determine the complexity of the processes described through the algorithms, an important quality when relating the characteristics of DMs to the thresholds defined at each decision node. The next section presents a decision theoretical model that incorporates the behavioral characteristics of DMs into the generation of the evaluation thresholds.

3. A simple information retrieval model

The structure of the benchmark algorithm is based on a series of factors that determine the probability of clicking on a given alternative according to its relative ranking position. Such a feature has been complemented by the introduction of a decision theoretical model that accommodates this type of behavior and relates it to the design of the algorithm.

The decision theoretical model considers explicitly the possibility of consulting websites to gain additional information when describing the threshold values that define the evaluation behavior of DMs through the algorithm. These threshold values are determined by the utility gain that may be obtained through the information retrieved from a given alternative relative to the potential utility derived from acquiring information on the next alternative ranked by the engine.

More precisely, the formal model defines the cutoff values as a function of the characteristics composing the alternatives, the subjective preferences of DMs, and their trust on the ranking results obtained. Note that the model provides a description of how to relate the cutoff values to the preferences of DMs. A geometric distribution could be used to determine the decreasing sequence of trust in the alternatives instead of the empirical CTRs obtained by Dean (2019). In this regard, the model and the algorithms are sufficiently flexible to incorporate any modification to the beliefs and preferences of DMs.

3.1. Formalization and technical assumptions

Assume that the alternatives are defined via tuples $(x_1, x_2, y_1, y_2) \in X_1^2 \times X_2^2$, which describe two different types of characteristics taking values in nonempty sets X_1 and X_2, with $x_1, x_2 \in X_1$ and $y_1, y_2 \in X_2$. In particular, the characteristics defining the alternatives can either be evaluated through the snippets, X_1, or require additional information retrievable from the corresponding links, X_2. This categorization constitutes an application of the search and experience attributes characterized by Gönsch (2020) when evaluating products.

An evaluation and decision model requires defining a *preference relation* $\geq$ on X_i, $i = 1, 2$, which is a reflexive, complete, and transitive binary relation on X_i. A utility function $u_i : X_i \to R$ representing $\geq$ *on* X_i preserves the order defined by the preference relation, satisfying $x' \geq x''$ $\Leftrightarrow u_i(x') \geq u_i(x'')$, $\forall x', x'' \in X_i$.

We identify X_1 and X_2 with closed real subintervals $X_1 = [x_i^m, x_i^M]$, $i = 1, 2$, and $X_2 = [y_j^m, y_j^M]$, $j = 1, 2$, such that $0 < x_i^m < x_i^M$ and $0 < y_j^m < y_j^M$. We introduce increasing and continuous additive utility functions denoted by $u : X_1^2 \times X_2^2 \to R$, such that $u(x_1, x_2, y_1, y_2) = u_1(x_1, x_2) + u_2(y_1, y_2)$, $\forall (x_1, x_2, y_1, y_2) \in X_1^2 \times X_2^2$, where u_k describes the subjective utility of the DM defined on the set X_k^2, $k = 1, 2$.

We also interpret X_i as a continuous random variable and assign a probability function $f_{X_i} : X_i \to [0, 1]$ to each set of characteristics. That is, DMs assign a probability $f_{X_i}(x_i)$ to the fact that the realization of the ith characteristic of an alternative that is evaluated randomly equals $x_i \in X_i$. The random variables will be assumed to be independent, though this assumption can be modified to allow for correlated characteristics.

Whenever utility functions are non-linear, the analysis must rely on the reference benchmark provided by the certainty equivalent value, $ce_i = u_i^{-1}(E_i)$, instead of the expected utility, E_i.

3.2. Convolution of two random variables

The next set of computations requires defining the convolution of two uniform random variables, which consists of the density assigned to their sum. That is, assume that $X_A, X_B \in$ R, both uniformly defined on the interval [0,1], have associated the following density

$$f_{X_A}(x) = f_{X_B}(y) = \begin{cases} 1, & \text{if} 0 \leq x,\ y \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

The sum of both variables $Z = X_A + X_B$ has the following density function

$$f_Z(z) = \int_{-\infty}^{\infty} f_{X_A}(z-y) f_{X_B}(y) dy$$

with $z = x + y$. Given the fact that the uniform densities are defined within the interval [0,1], the previous density function simplifies to

$$f_Z(z) = \int_0^1 f(z-y) dy$$

with

$$f(z-y) = \begin{cases} 1, & \text{if} \quad 0 \leq z-y \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

implying that $z - 1 \leq y \leq z$. The integrals defining the convolution are determined by the interval scenarios $0 < z < 1$ and $1 < z < 2$, defined for two uniform random variables. The integration limits follow from z - l < y < z and the fact that y is defined through the z interval described above. The subsequent equation of the convolution is given by

$$f_Z(z) = \begin{cases} \int_0^z dy, & \text{if} \quad 0 \leq z \leq 1 \\ \int_{z-1}^1 dy, & \text{if} \quad 1 \leq z \leq 2 \\ 0, & \text{otherwise} \end{cases}$$

which simplifies to

$$f_Z(z) = \begin{cases} z, & \text{if} \quad 0 \leq z \leq 1 \\ 2-z, & \text{if} \quad 1 \leq z \leq 2 \\ 0, & \text{otherwise} \end{cases}$$

This density allows us to work within a two-dimensional setting that defines a unique crossing point between the functions determining the retrieval behavior of DMs. The next subsection illustrates how this crossing point can be derived from a formal decision theoretical environment designed according to the standard principles of expected utility theory.

3.3. Retrieval utilities

The functions representing the expected utilities that DMs derive from their information retrieval processes formalize their behavior when reading the snippets and evaluating a given alternative, J. That is, after evaluating the initial characteristics of an alternative, the DM must decide between clicking on the link and retrieving further information about the alternative or moving on to the next alternative in the ranking. The corresponding decision is based on the values $x_1, x_2 \in X_1$ observed for J and, therefore, conditioned by the convolution of the characteristics defining the different alternatives.

We have simplified the presentation assuming alternatives composed by two characteristics within each set X_1 and X_2. This assumption requires defining the convolution of the densities assigned to the characteristics together with a subjective utility based on the relative distance between each characteristic and the most preferred evaluation subjectively defined by DMs.

The initial utility function focuses on the characteristics of the alternatives described in the snippets, $x_i \in X_1$, $i = 1, 2$, and their value relative to the best potential realization considered by the DM, x_i^M. Note that the DM must also account for the set of realizations of the second characteristic, $y_j \in [y_j^m, y_j^M]$, $j = 1, 2$, with $y_j \in X_2$, that may be observed when evaluating the alternative further. The corresponding expected utility is therefore defined as follows

$$U(z_1, z_2) \stackrel{def}{=} u_1(z_1) + \int_{z_2 \in Z_2} \mu_2(z_2)(u_2(z_2))\, dz_2 \tag{1}$$

with

$$u_1(z_1) = \frac{1}{i} \sum_i \left(\frac{x_i - x_i^m}{x_i^M - x_i^m} \right), \quad x_i \in X_1 \tag{2}$$

$$E_2 = \int_{z_2 \in Z_2} \mu_2(z_2)(u_2(z_2))\, dz_2 = \frac{1}{j} \int_{z_2 \in Z_2} f(z_2) \sum_j \left(\frac{y_j - y_j^m}{y_j^M - y_j^m} \right) dz_2, \quad y_j \in X_2 \tag{3}$$

Several basic remarks follow.

- In order to keep notation consistent, we define $z_1 = \sum_{i=1}^{2}\left(\frac{x_i - x_i^m}{x_i^M - x_i^m}\right)$ and $z_2 = \sum_{j=1}^{2}\left(\frac{y_j - y_j^m}{y_j^M - y_j^m}\right)$, implying that there are two different sets of double characteristics, one directly observable, the other requiring a click on the corresponding link. Triples could be defined per characteristic set without increasing significantly the complexity of the presentation, but the main qualitative results would remain unchanged.
- We have assumed risk neutral DMs endowed with linear utility functions to formalize the evaluation metric within each characteristic. This assumption could be easily relaxed, and the shape of the utility functions generalized, making use of the certainty equivalent values in place of the expected utilities. However, this extension would simply complicate the presentation without modifying the main intuition provided or qualitative results obtained.
- The uncertainty faced by DMs when evaluating the characteristics is formalized through a uniform density of potential realizations, which are distributed within $[x_i^m, x_i^M]$ and $[y_j^m, y_j^M]$ for $i,j = 1,2$. The uniform probability distribution reflects the highest information entropy faced by DMs regarding the results obtained from the search engine (Di Caprio et al., 2022b; Tavana et al., 2015b).

The expected utility derived from evaluating the initial set of characteristics from one of the alternatives composing the ranking is defined as follows

$$U(z_1, z_2) \stackrel{def}{=} \frac{1}{i}\sum_i\left(\frac{x_i - x_i^m}{x_i^M - x_i^m}\right) + \frac{1}{j}\int_{z_2\in Z_2} f(z_2)\sum_j\left(\frac{y_j - y_j^m}{y_j^M - y_j^m}\right)dz_2 \quad (4)$$

Similarly, the expected utility obtained from a random evaluation of one of the alternatives within the ranking equals

$$C(z_1, z_2, \Phi) \stackrel{def}{=} \int_{z_1\in Z_1} \mu_1(z_1)(u_1(z_1))\,dz_1 + \Phi\int_{z_2\in Z_2} \mu_2(z_2)(u_2(z_2))\,dz_2 \quad (5)$$

Adapting Eq. (3) and substituting the corresponding expression within Eq. (5) we obtain an extended equation directly comparable with $U(z_1, z_2)$

$$C(z_1, z_2, \Phi) \stackrel{def}{=} \frac{1}{i}\int_{z_1\in Z_1} f(z_1)\sum_i\left(\frac{x_i - x_i^m}{x_i^M - x_i^m}\right)dz_1 + \frac{\Phi}{j}\int_{z_2\in Z_2} f(z_2)\sum_j\left(\frac{y_j - y_j^m}{y_j^M - y_j^m}\right)dz_2 \quad (6)$$

where $\Phi \in [1,2]$ has been introduced as a compensation mechanism to account for the uncertainty inherent to the characteristics defining the alternatives, a subset of which remains unknown through the initial evaluation process (Herrmann, 2015). The trust placed by DMs on the rankings delivered by the search engines implies that Φ can be assumed to increase with the alternatives located in lower ranking positions.

Consider the case where the individual characteristics composing X_1 and X_2 are uniformly distributed within the interval $[0,10]$, leading to a normalized expected value of 0.5 for both z_1 and z_2. It therefore follows that $U(z_1, z_2)$ is a linear function defined within $[0.5,1.5]$, and $C(z_1, z_2, \Phi) \in [1,1.5]$, the value of $\Phi \in [1,2]$ depending on the trust placed on the corresponding alternative being able to fulfill the subjective preferences of the DM.

We have designed this formal environment to identify the value of $\Phi \in [1,2]$ such that $U(z_1, z_2)$ and $C(z_1, z_2, \Phi)$ cross at the value of z_1 generating the empirical CTRs observed for each alternative based on their positions within the ranking. That is, the value of Φ subjectively defined by DMs generates the value of z_1 that determines their retrieval behavior per alternative. We could therefore interpret $(2 - \Phi)$ as the subjective degree of trust assigned to each alternative based on its relative ranking position.

Consider the convolution of two initial characteristics, $x_i \in X_1, i = 1, 2$, and the largest and smallest CTRs observed empirically by Dean (2019) – described in Fig. 1 and, more specifically, the second column of Table 3.

- The value of Φ required to obtain a z_1 cutoff such that $f_Z(z_1) = 0.68$ is derived as follows. The value of z_1 delivering the probability mass required is given by $z_1 = 1.2$. We normalize this value within Fig. 1 to $z_1 = 0.6$ and solve $U(z_1, z_2) = 0.6 + 0.5 = 0.5 + 0.5\Phi = C(z_1, z_2, \Phi)$ for $\Phi = 1.2$.
- The same intuition applies when $f_Z(z_1) = 0.97$, which is solved for $z_1 = 1.7551$. Its normalized value, $z_1 = 0.8776$, is used to find the value of $\Phi = 1.7552$ solving $1.3776 = 0.5 + 0.5\Phi$.

Fig. 2 illustrates both thresholds as defined by $U(z_1, z_2)$ and $C(z_1, z_2, \Phi)$, together with the basic reference case for $\Phi = 1$.

We must emphasize the fact that the model has been defined to complement the design of the retrieval algorithms and provide additional intuition when studying the online behavior of users. That is, the model is not an exhaustive analysis of the generation of evaluation thresholds and the corresponding retrieval probabilities, but an intuitive formalization of the complexities involved within seemingly simple decision processes.

The algorithmic evaluation scenarios defined allow for the introduction of signals and modifications in the beliefs of DMs as the retrieval process develops. This possibility emphasizes the complexity inherent to the search and evaluation processes defined by DMs. The heuristic mechanism validated numerically does not imply simplistic behavior on the side of DMs but defines a selective feature allowing to shortlist the alternatives closest to their preferences.

4. On the complexity of evaluation processes

The empirical analyses of online information retrieval behavior describe users who follow sequential decision-tree processes (Hendahewa & Shah, 2017). However, researchers seem reticent to formalize the subsequent decision-tree algorithmic structures. This lag in the literature may be due to the belief that the incorporation of heuristic mechanisms following from bounded rationality assumptions results in trivial decision-making environments. This would be the case if we were to simulate ten independent random realizations, assign one to each alternative and omit any incentives resulting from the observations

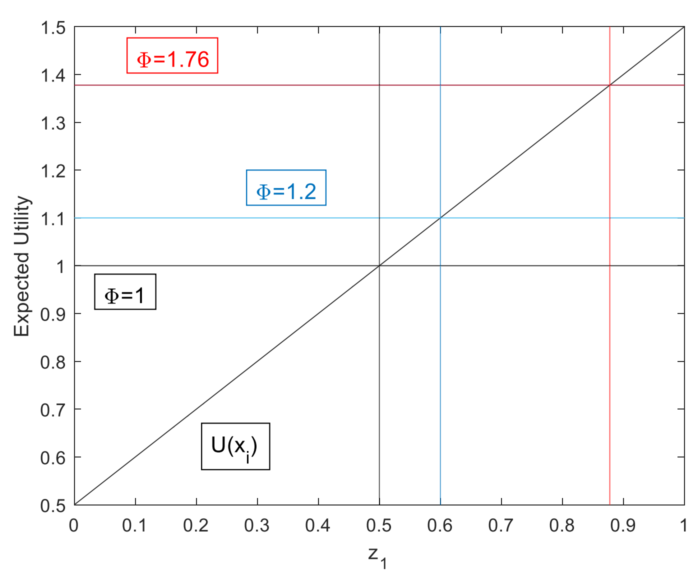

Fig. 2. Evaluation thresholds determined by the main (Φ, z_1) reference pairs.

retrieved. Such an assumption would imply that DMs do not follow any predetermined search strategy but click randomly on pages independently of the results obtained through the sequential search process.

More precisely, consider the definition of the main variable used to analyze the behavior of DMs when retrieving information online. The CTRs obtained from a given number of search queries, n, are defined as follows

$$\text{CTR of alternative } i = \frac{\text{Number of users clicking on the link to alternative } i}{\text{Number of users performing a search}}, \; i = 1, \ldots, n.$$

Assume that a DM only checks the first two alternatives displayed by the search engine. The resulting string of data describing the behavior of the DM would consist of vectors with ten entries where the first two are different from the remaining ones, which could be assigned a value of zero or any predetermined default one for analytical purposes. Consider now the behavior of a DM who sets out to select two satisficing alternatives among those displayed within the first page of results described by the engine.

The retrieval process of the second DM is considerably more complex than that of the first one. Indeed, the second DM will evaluate the alternatives in the order displayed by the engine and select the first two that satisfy the predetermined requirement. Clearly, these alternatives are not necessarily the ones located in the first two positions within the ranking and could be located anywhere within it. That is, the fact that DMs trust the ranking displayed by the engine does not mean that they follow it blindly, simply that they proceed through the results in the order displayed by the engine and then apply their corresponding selection criteria when deciding which alternatives to click.

The above intuition implies that a basic retrieval algorithm mimicking the CTRs observed empirically can be easily defined. Each run of the algorithm is naturally considered to represent a search query by a user. Thus, given the definition of CTR, a basic algorithm such as those presented in Figs. 3 and 4 – but extended to incorporate ten independent evaluations – would suffice to generate the CTRs observed empirically. We simply need to assume that DMs assign a predetermined probability to clicking on each alternative. Note, however, that the retrieval process and the algorithm lack any interaction across observations. That is, assume that the DM evaluates two alternatives whose characteristics are expected to perform above a given predetermined threshold. The corresponding search process is limited to any two alternatives selected ex ante by the DM from the initial ranking. As a result, the behavior of the DM remains unaltered by the actual evaluations observed.

Assume now that the DM aims at observing two alternatives that display satisficing evaluations. Accounting for this possibility requires defining an algorithm that considers all the potential paths that may be followed by the DM. The corresponding algorithm should account for an increasingly complex decision tree incorporating all the binary decision nodes that may arise from the information retrieval process of DMs (Schulz, 2008).

Table 1 describes the main differences – in terms of decision nodes – among the algorithms required to simulate the basic retrieval process with independent evaluations, a standard binary tree, and the complete retrieval process where DMs aim at evaluating a given number of satisficing alternatives. The binary tree and the complete evaluation scenario define the same retrieval process when considering the whole set of alternatives composing the first page of results. In this case, the decision tree defined to replicate the CTR behavior of DMs is composed by $2^{10} - 1 = 1023$ binary decision nodes and has a total of $\left(\sum_{n=1}^{10} 2^n\right) + 1 = 2047$ nodes.

Thus, even when implementing a heuristic mechanism where DMs aim at finding two satisficing alternatives out of a total of ten, each characteristic set composed by two independent random variables, the subsequent retrieval process requires considering 55 potential binary decisions.

4.1. Decision trees and satisficing algorithms

The stochastic sequential evaluation structure of the algorithm is built as follows. We assign a random uniform realization within [0, 1] to each alternative as DMs proceed through the ranking displayed by the engine. Each alternative is also assigned a cutoff value, $c_i, \; i = 1, \ldots, 10$, based on its position within the ranking. The behavior of DMs is determined by the value of each realization relative to the cutoff assigned to the corresponding alternative.

That is, when considering the first alternative composing the ranking, the DM either clicks on the link provided or proceeds with the

```
nrows = 20;
ncols = 1000000;
A = zeros(nrows,ncols);

for i = 1
   for j = 1:ncols
        A(i,j)=rand(1);
        if A(i,j)>0.68
        A(11,j)=1;
        end
          A(i+1,j)=rand(1);
          if A(i+1,j)>0.75
          A(12,j)=2;
          end
    end
end
```

Fig. 3. Code of the basic algorithm evaluating two alternatives but lacking interactions across variables and evaluations.

```
nrows = 20;
ncols = 1000000;
A = zeros(nrows,ncols);

for i = 1
   for j = 1:ncols
        A(i,j)=rand(1);
        if A(i,j)>0.68
        A(11,j)=1;
        end
          A(i+1,j)=rand(1);
          if A(i+1,j)>0.75
          A(12,j)=2;
          end
            A(i+2,j)=rand(1);
            if A(i+2,j)>0.81
            A(13,j)=3;
            end
   end
end
```

Fig. 4. Code of the basic algorithm evaluating three alternatives but lacking interactions across variables and evaluations.

Table 1
Complexity inherent to the different types of retrieval processes.

	Number of alternatives	1	2	3	4	5	6	7	8	9	10
Binary Tree	Binary decision nodes	1	3	7	15	31	63	127	255	511	1023
	Number of final nodes	2	4	8	16	32	64	128	256	512	1024
	Total number of nodes	3	7	15	31	63	127	255	511	1023	2047
Independent	Binary decision nodes	1	2	3	4	5	6	7	8	9	10
	Number of final nodes	2	4	6	8	10	12	14	16	18	20
	Total number of nodes	3	6	9	12	15	18	21	24	27	30
Complete	Binary decision nodes	10	55	175	385	637	847	967	1012	1022	1023
	Number of final nodes	11	56	176	386	638	848	968	1013	1023	1024
	Total number of nodes	21	111	351	771	1275	1695	1935	2025	2045	2047

second alternative. The decision is based on the alignment between the preferences of the DM and the information displayed, namely, clicks result from the realization being higher than the cutoff value. The remaining part of the sequential process is defined similarly, with the DM considering two paths per potential decision node. The retrieval process is therefore defined by a sequential decision-tree structure of binary choices whose complexity increases as the DM proceeds through the ranking.

We design a series of satisficing heuristic algorithms that interrupt the information retrieval process when DMs observe a predetermined number of alternatives aligning with their subjective preferences. The basic satisficing algorithm assumes that DMs conclude the retrieval process after observing one alternative aligning with their preferences. Further algorithms are defined requiring DMs to observe two, three and ten alternatives aligning with their preferences. Figs. 1–3 describe the information retrieval processes followed by DMs when considering one, two and three satisficing alternatives before concluding their search, respectively.

For instance, Fig. 6 illustrates the evaluation scenario where DMs aim at observing two satisficing alternatives through their information retrieval processes. The intuition describing the process follows from the information retrieval behavior of a DM who observes and evaluates alternatives in the order provided by the search engine. The retrieval process ends after the DM observes two alternatives aligning with his preferences that deliver a sufficient – satisficing – level of utility. As explained in Section 3, this level is determined subjectively by each DM, who will conclude the search after retrieving information from the first ten alternatives provided by the engine even if he has been unable to identify two satisficing ones.

Consider once again the information retrieval process described in Fig. 6. Note that, if the first two alternatives deliver the satisficing utility requested, the retrieval process ends with the DM having clicked on the first and second alternatives. Thus, if all DMs apply the same behavioral rule based on the assumption that, given their ranking positions, the first two alternatives deliver the highest potential utility, we should observe all DMs clicking on the first two alternatives and ending their searches right after.

The behavior observed would therefore consist in all DMs clicking on the first two alternatives, each one of them receiving half the clicks and accounting for half the average traffic share. The CTRs would be equal to 100 % for each alternative, since they are both clicked by DMs on each query search. However, this is not what can be observed in the data (Chitika, 2013; Dean, 2019), with DMs retrieving information from the whole set of alternatives composing the first page of results provided by the engine. In this regard, Fig. 6 describes the potential scenarios that may arise as any of the first two alternatives, or both, fail to deliver the utility required by the DMs. Figs. 5 and 7 describe similar retrieval settings, with DMs aiming for a total of one or three satisficing alternatives being observed throughout their evaluation processes, respectively.

In order to provide additional intuition, Fig. 8 presents a simplified version of the codified structure of the algorithm accounting for three satisficing alternatives, that is, ending the retrieval process when the DM observes three alternatives aligning with his preferences. As already

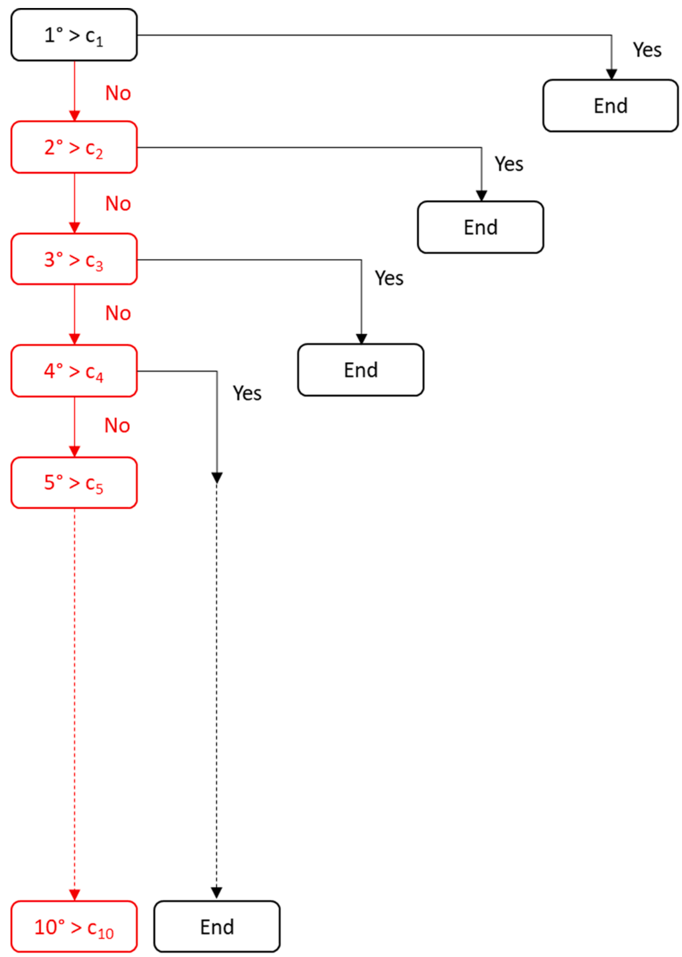

Fig. 5. Satisficing heuristic framework with one satisficing alternative.

emphasized, the complexity of the heuristic algorithms ranges from the simplest one – requiring a unique satisficing alternative and composed by ten decision nodes – to the $(2^{10} - 1)$ nodes defining the algorithm that requires ten satisficing alternatives, namely, the binary evaluation of all the alternatives ranked within the first page of results delivered by the search engine.

Tables A2 and A3 within the supplementary appendix section provide the MATLAB codes used to simulate the satisficing frameworks with two and three alternatives, respectively. The increase in the complexity of the code is evident, as well as the subsequent imposition on the computational abilities of DMs. As explained in the previous subsection, Figs. 3 and 4 present the algorithms corresponding to the independent retrieval processes focusing on evaluating two and three alternatives, respectively. Their relative simplicity becomes particularly evident when compared to the algorithms presented in the appendix sections.

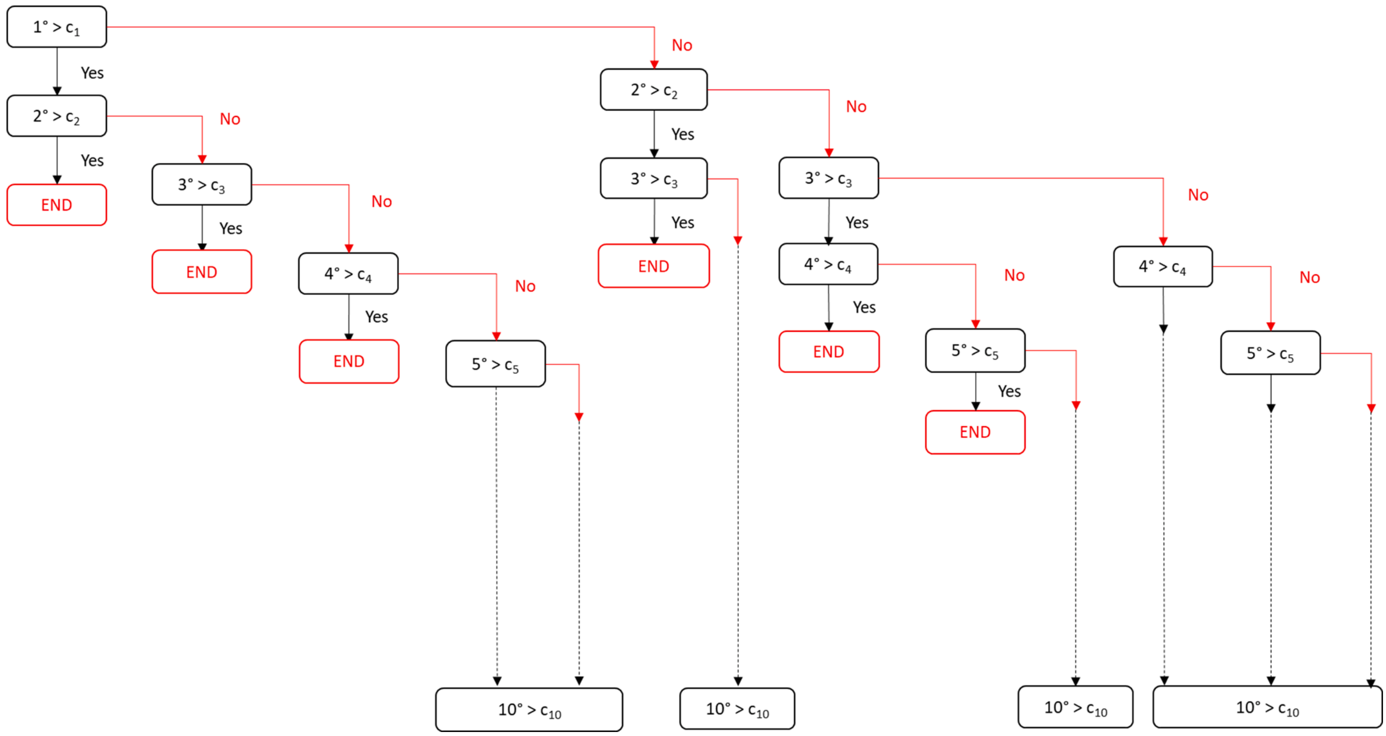

Fig. 6. Satisficing heuristic framework with two satisficing alternatives.

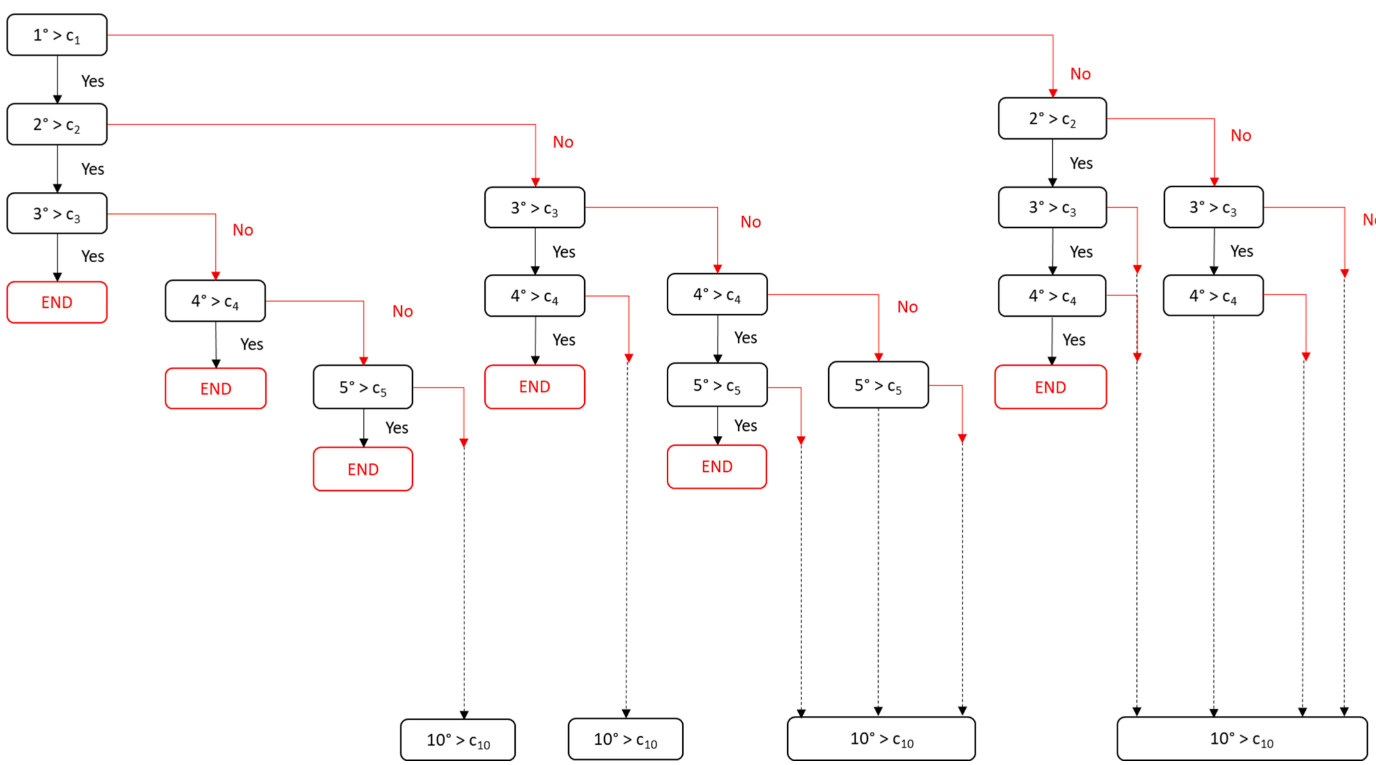

Fig. 7. Satisficing heuristic framework with three satisficing alternatives.

The subsequent tables composing the appendix illustrate the substantial increment in the complexity of the retrieval processes assumed on the DMs as a higher number of alternatives are incorporated into their satisficing objectives.

We run several simulations illustrating the performance of the algorithms and their main differences relative to a basic retrieval framework delivering similar CTR results. The simulations have been designed to highlight the actual complexity involved in the formalization of the information retrieval process when considering the whole set of interactions across evaluations and alternatives.

The set of benchmark algorithms is sufficiently flexible to account for the effects of strategic signals received at any point throughout the

```
generate random uniform_i
          if random uniform_i > c_i
          page is visited and evaluated;
          random uniform_{i+1};

                    if random uniform_{i+1} > c_{i+1}
                    page is visited and evaluated;
                    random uniform_{i+2};

                              if random uniform_{i+2} > c_{i+2}
                              page is visited and evaluated;

                              else
                              random uniform_{i+3};
                              if random uniform_{i+3} > c_{i+3}
                              page is visited and evaluated;
                              ….
                              end

                    else
                    if random uniform_{i+2} > c_{i+2}
                    page is visited and evaluated;
                    random uniform_{i+3};

                              if random uniform_{i+3} > c_{i+3}
                              page is visited and evaluated;

                              else
                              random uniform_{i+4};
                              if random uniform_{i+4} > c_{i+4}
                              page is visited and evaluated;
                              ….
                              end

                    else
                    if random uniform_{i+3} > c_{i+3}
                    page is visited and evaluated;
                    random uniform_{i+4};
                    ….
                    end

          else
          if random uniform_{i+1} > c_{i+1}
          page is visited and evaluated;
          random uniform_{i+2};

               if random uniform_{i+2} > c_{i+2}
               page is visited and evaluated;
               random uniform_{i+3};

                         if random uniform_{i+3} > c_{i+3}
                         page is visited and evaluated;

                         else
                         random uniform_{i+4};
                         if random uniform_{i+4} > c_{i+4}
                         page is visited and evaluated;
                         ….
                         end

               else
               if random uniform_{i+3} > c_{i+3}
               page is visited and evaluated;
               random uniform_{i+4};
               ….
               end

          else
          if random uniform_{i+2} > c_{i+2}
          page is visited and evaluated;
          random uniform_{i+3};
          ….
          end

end
```

Fig. 8. Summary of the codified structure of the algorithm requiring three satisficing alternatives.

information retrieval process. Any type of behavioral modification can be incorporated into the analysis to illustrate its expected effect on the set of resulting CTRs. This feature constitutes a significant advantage over the simpler algorithms that do not account for the set of potential interactions across evaluations and alternatives arising through the sequential information retrieval process of DMs.

4.2. *Categorizing evaluation processes*

The next result provides important intuition regarding the categorization difficulties that follow from the different retrieval processes generated. That is, the basic algorithm defining DMs that perform independent searches and the complete one – where DMs search for a given number of satisficing alternatives – converge to the same stochastic framework and produce identical CTRs.

Consider a basic retrieval process based on the realizations from ten independent random variables uniformly distributed in [0, 1], where each alternative is assigned a cutoff value within this interval determining the evaluation behavior of DMs. The law of large numbers implies that simulating a sufficiently large number of realizations would indeed lead to a percentage of clicks identical to the threshold values defined. The complete algorithm, where DMs evaluate ten alternatives out of a total of ten, leads to the same retrieval structure. That is, DMs will consider all ten alternatives per query, leading to a stochastic structure whose CTRs converge to those defined by ten independent and unrelated trials. However, as can be intuitively understood, the retrieval behavior is substantially different in both cases, the latter being closer to the actual rational behavior considered by economists and decision theorists, while the former lacks a structured search strategy.

Proposition. The basic algorithm – defining the independent evaluation of n alternatives – and the complete one, where DMs aim to observe n satisficing alternatives from a given set, deliver the same CTRs when DMs set out to observe n satisficing alternatives out of a total of n.

Proof. Consider the retrieval setting with ten alternatives. The proof relies on noting that, when aiming to observe ten satisficing alternatives, the probability of evaluating each one of them equals one, leading to the same stochastic framework as the basic algorithm, where ten alternatives are evaluated with probability one. The observations retrieved determine the clicking decisions when compared to the corresponding threshold values. ■

Thus, while generating CTRs identical to those observed in real life environments is an almost trivial problem, incorporating the set of potential interactions across evaluations and alternatives requires defining a much more complex algorithmic structure. This result provides valuable intuition when considering search processes based on less than ten satisficing alternatives, since, as will be illustrated in the next section, the complete algorithm delivers the required CTRs when DMs set out to observe three satisficing alternatives out of a total of ten.

5. Numerical analysis and main findings

The algorithms have been calibrated using the values provided by Dean (2019), who analyzed five million queries to derive the CTRs on the organic alternatives ranked by Google within the first page of results. The cutoff values assigned to each node of the tree correspond to the observed CTRs, which, therefore, define the probability of clicking on an alternative. The CTRs obtained by Dean (2019) are presented in the second column of Table 3, while the corresponding set of cutoff values is summarized in the first column of Table 2 and given by

$[c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8, c_9, c_{10}] = [0.68, 0.75, 0.81, 0.86, 0.90, 0.94, 0.96, 0.97, 0.97, 0.97]$

Table 2 describes intuitively the behavior of DMs. In particular, this table presents realizations from the different algorithmic scenarios simulated, each of its columns representing the search query results obtained by a DM and his subsequent information retrieval behavior.

It may seem naturally plausible to assume that DMs stop retrieving

Table 2
Simulation samples from the different satisficing scenarios.

Cutoff Values		Number of Satisficing Alternatives															
		One				Two				Three				Ten			
0.68	**Stochastic Realizations**	0.533	0.876	0.550	0.171	0.853	0.376	0.714	0.889	0.891	0.428	0.963	0.738	0.521	0.788	0.787	0.174
0.75		0.351	–	0.622	0.228	0.482	0.239	0.102	0.017	0.953	0.759	0.297	0.508	0.022	0.983	0.925	0.956
0.81		0.939	–	0.587	0.436	0.272	0.243	0.447	0.148	0.181	0.923	0.710	0.865	0.599	0.179	0.928	0.985
0.86		–	–	0.208	0.311	0.731	0.019	0.693	0.959	0.074	0.872	0.901	0.140	0.338	0.433	0.990	0.941
0.90		–	–	0.301	0.923	0.969	0.983	0.333	–	0.988	–	0.874	0.657	0.784	0.416	0.920	0.301
0.94		–	–	0.471	–	–	0.803	0.922	–	–	–	0.829	0.593	0.896	0.964	0.431	0.927
0.96		–	–	0.230	–	–	0.563	0.632	–	–	–	0.035	0.513	0.426	0.108	0.778	0.531
0.97		–	–	0.844	–	–	0.389	0.932	–	–	–	0.906	0.245	0.998	0.811	0.419	0.860
0.97		–	–	0.195	–	–	0.653	0.426	–	–	–	0.628	0.974	0.831	0.460	0.664	0.273
0.97		–	–	0.226	–	–	0.994	0.300	–	–	–	0.573	–	0.629	0.626	0.195	0.943
Pages Clicked	3	1	–	5	1	5	1	1	1	2	1	1	8	1	1	2	
	–	–	–	–	5	10	–	4	2	3	4	3	–	2	2	3	
	–	–	–	–	–	–	–	–	5	4	–	9	–	6	3	4	
	–	–	–	–	–	–	–	–	–	–	–	–	–	–	4	–	
	–	–	–	–	–	–	–	–	–	–	–	–	–	–	5	–	
	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	
	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	
	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	
	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	
	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	

Table 3
CTRs differences across satisficing scenarios.

		Number of Satisficing Alternatives									
Ranking position	Dean (2019)	One	Two	Three	Four	Five	Six	Seven	Eight	Nine	Ten
1	31.7	31.93	32.04	32.02	32.01	32.04	32.02	32.00	31.90	32.00	32.02
2	24.7	17.01	24.98	24.99	24.97	25.03	24.96	24.96	25.01	24.91	24.98
3	18.7	9.67	17.44	18.97	18.98	19.00	19.01	18.97	19.02	19.01	19.03
4	13.6	5.76	11.79	13.83	13.96	14.01	14.03	13.99	13.96	13.95	13.98
5	9.5	3.55	7.85	9.64	9.97	10.04	10.03	10.05	10.04	9.98	9.99
6	6.2	1.93	4.45	5.69	5.99	6.01	6.02	5.98	5.99	6.02	6.02
7	4.1	1.20	2.85	3.75	3.97	3.97	3.98	4.02	4.02	4.00	3.99
8	3.1	0.88	2.10	2.79	2.96	2.98	3.00	3.04	3.03	2.94	2.98
9	3	0.83	2.04	2.77	2.95	3.00	3.00	3.02	2.98	2.99	3.00
10	3	0.82	2.00	2.76	2.94	3.01	3.00	2.99	3.01	3.01	3.00

information after observing one alternative aligning with their preferences. The third column of Table 3 has been introduced to refute this type of reasoning. The intuition behind this result is straightforward and follows from the differences between CTRs and the distribution of probability through the nodes composing the retrieval process, which binds the sum of clicks on the alternatives to a value of one. This constraint is however inherent to the structure of the basic algorithm considering a unique satisficing alternative.

The remaining columns of Table 3 compare the CTRs obtained by Dean (2019) with those delivered by the heuristic algorithms requiring two, three and ten satisficing alternatives. We have executed one million runs of the algorithms per information retrieval scenario. A noticeable result is the fact that the performance of the algorithm with two satisficing alternatives constitutes a sufficient approximation to the CTRs observed, while adding a third alternative provides an almost identical set of CTRs. That is, the heuristic algorithms have the capacity to replicate the behavior observed when accounting for three satisficing alternatives. The mimicking capacity of the algorithms displays marginal improvements as we increase the number of satisficing alternatives, though it is clearly not necessary to consider the 2047 nodes composing the algorithm with ten alternatives to simulate the actual behavior of DMs.

Fig. 9 presents the retrieval profiles generated through increasingly complex evaluation scenarios. Note how two satisficing alternatives already deliver a retrieval framework similar to the limit case with ten alternatives. Thus, as already illustrated, we do not require overly complex retrieval frameworks or DMs endowed with highly enhanced assimilation capacities to generate the CTRs observed empirically. Each scenario within the figure accounts for 300 queries, each of them amounting for up to 10 realizations, depending on the scenario analyzed and the value of the random realizations observed. Whenever an alternative is not evaluated or a realization does not lead to a click, a value of zero is assigned to the corresponding matrix entry.

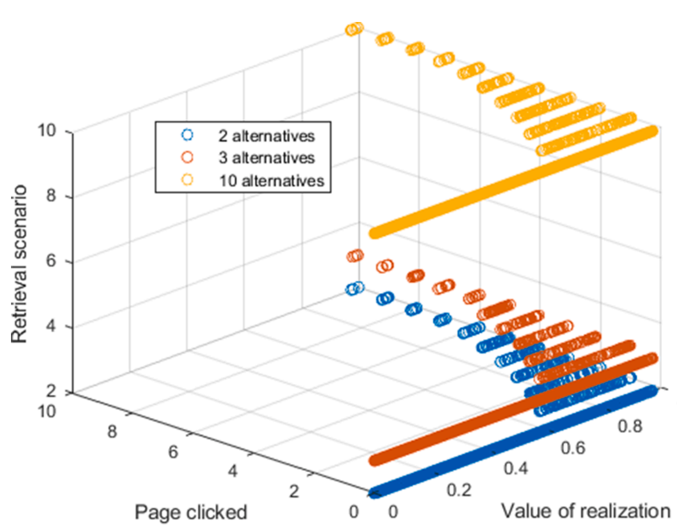

Fig. 9. Retrieval profiles corresponding to increasingly complex retrieval scenarios.

A more concise illustration of the intuition described in Fig. 9 is provided in Fig. 10, which describes the absolute value of the difference between the CTRs reported by Dean (2019) and those obtained from one million runs of the benchmark algorithms per information retrieval scenario. Note how the significant differences observed in the scenario with one alternative are quickly smoothed as a second and third alternative are incorporated by the DM into the corresponding evaluation processes. We conclude the analysis by noting that these results are in accordance with the empirical evidence presented by Jansen et al. (1998) and Baeza-Yates (2005), who found that users tend to click an average of two pages per search query.

As stated at the beginning of this section, the numerical simulations performed are based on CTRs retrieved from organic search outcomes delivered by Google. We must note that the results presented are not constrained to this type of search engine but apply to any search process where DMs must evaluate a set of characteristics and decide whether they want to click on the corresponding link or prefer to proceed with the next alternative. This would be the case when considering vertical search engines such as TripAdvisor.

That is, CTRs have been translated into clicking probabilities that define the cutoff points corresponding to each alternative according to its position within the ranking – and subsequent decision tree –. However, these points could be endogenously defined as part of the decision retrieval model presented in Section 3. In other words, this model introduces a basic decision setting defined to illustrate how the cutoff values can be related to the formal information retrieval and evaluation criteria of DMs. The model could be extended to consider explicitly the interactions among the characteristics defining the different alternatives and their evaluations, together with the potential realizations of the characteristics defining the alternatives remaining to be observed (Tavana et al., 2016a, 2016b). This type of retrieval framework would allow to modify specific characteristics of each alternative – conditioned by their relative importance to DMs – and analyze the resulting evaluation scenarios, widening the set of strategic signaling environments that could be defined by the firms or economic agents whose products are ranked by the engine.

6. Discussion: on retrieval processes and assimilation capacities

The results described in the previous section highlight an important problem from a modelling viewpoint with implications regarding the information assimilation capacities of DMs and their subsequent formalization in decision sciences. The results shed light on the possibility that DMs do not follow a sequential structured search – where the information retrieved conditions the subsequent evaluation paths – but perform a simple search determined by a set of independent pre-determined cutoff values.

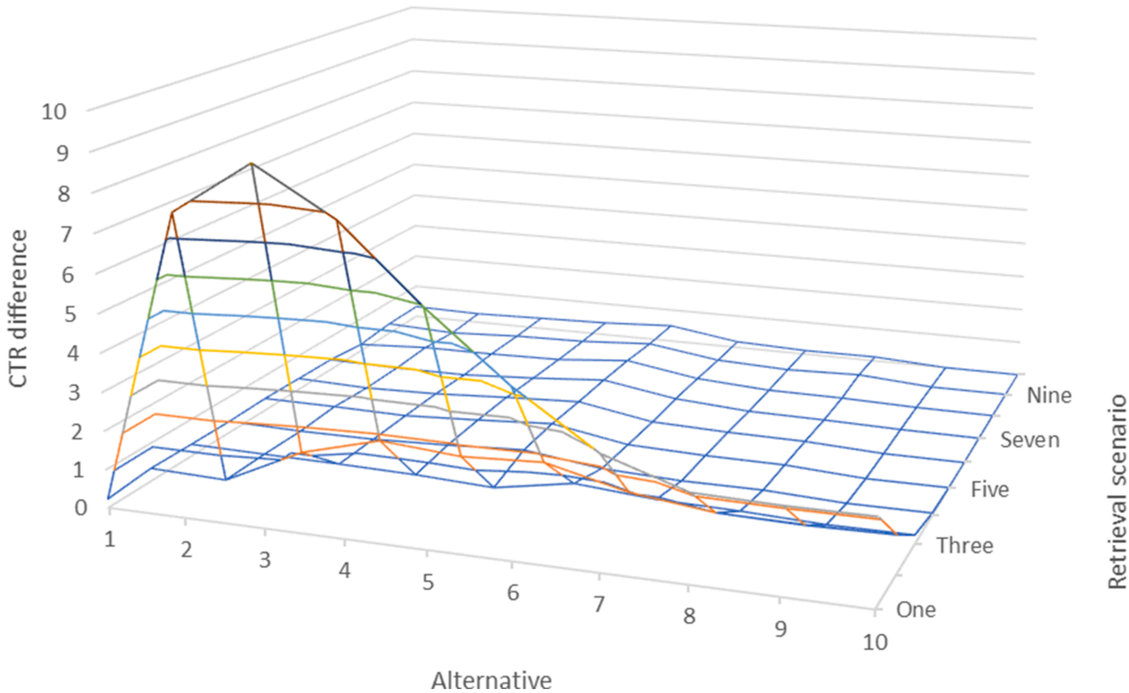

Fig. 10. Absolute value of the difference between the CTRs reported by Dean (2019) and those obtained from one million runs of the benchmark algorithms per retrieval scenario.

Table 4
CTR differences between retrieval processes with a 0.5 threshold value per alternative.

	Number of Satisficing Alternatives							
	Basic				Benchmark			
Ranking position	One	Two	Three	Ten	One	Two	Three	Ten
1	50.07	49.97	50.04	49.93	50.00	49.99	50.08	50.00
2	–	50.01	49.98	49.99	25.03	49.96	49.96	50.04
3	–	–	49.99	49.96	12.45	37.45	50.11	49.98
4	–	–	–	50.00	6.26	24.98	43.74	50.17
5	–	–	–	49.91	3.13	15.63	34.42	49.98
6	–	–	–	49.95	1.58	9.40	24.96	49.97
7	–	–	–	49.99	0.78	5.49	17.13	50.00
8	–	–	–	49.90	0.39	3.14	11.34	49.89
9	–	–	–	50.04	0.20	1.78	7.23	49.93
10	–	–	–	50.08	0.09	0.98	4.45	49.99

Given the fact that DMs tend to click on two pages per search query (Baeza-Yates, 2005; Jansen et al., 1998), assuming two independent observations would clearly not generate the CTRs derived from real-life scenarios but concentrate them on the first two alternatives composing the ranking. A comparison of the basic and complete algorithms illustrating this feature – with a 0.5 threshold value assigned to each alternative – is presented in Table 4, where the mimicking capacities of the corresponding retrieval frameworks can be compared. It seems intuitively plausible to conclude that the retrieval behavior of DMs is not based on basic evaluation processes, a conclusion reinforced when considering less than ten satisficing observations.

That is, information retrieval processes cannot be composed by a series of independent evaluations. DMs must consider – to some extent – the interactions across observations and alternatives arising from the set of potential paths that may be defined when proceeding through the ranking. In this way, we can reconcile the value of the CTRs observed in real-life environments with a sequential retrieval process where DMs aim at observing a given number of predetermined alternatives.

Besides illustrating the rationality inherent to these processes, the benchmark algorithms allow to analyze the consequences derived from introducing external modifications to the retrieval strategies of DMs. They also highlight the different assimilation capacities defining the behavior of DMs and whether differences across retrieval strategies can be identified and categorized accordingly. The set of evaluations considered by DMs and conditioning their retrieval behavior includes the different characteristics defining each alternative together with the potential observations that may be retrieved.

We address the question of how many satisficing alternatives must be considered by DMs to generate the CTRs observed in real-life settings. As illustrated in Table 3, a retrieval algorithm calibrated to the empirical CTRs delivers the required values when DMs aim at observing three satisficing alternatives, with two alternatives constituting a sufficient approximation. The correlation analysis presented in Table 5 demonstrates how the CTRs obtained are identical when considering three or more alternatives and sufficiently close when accounting for two. That is, we can validate the intuition that information retrieval strategies are neither trivial nor require substantial assimilation capacities on the side of DMs.

The search processes generated display information retrieval strategies that mimic the data available to computer scientists when categorizing the corresponding behavior through machine learning techniques (Qin et al., 2020). The algorithms deliver two different

Table 5
Pearson correlation between the CTRs obtained by Dean (2019) and the benchmark algorithms per retrieval scenario.

		Dean	One	Two	Three	Four	Five	Six	Seven	Eight	Nine	Ten
Dean	Pearson Correlation	1	.954**	.999**	1.000**	1.000**	1.000**	1.000**	1.000**	1.000**	1.000**	1.000**
	Sig. (2-tailed)		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	10	10	10	10	10	10	10	10	10	10	10
One	Pearson Correlation	.954**	1	.963**	.951**	.951**	.951**	.951**	.952**	.951**	.952**	.951**
	Sig. (2-tailed)	.000		.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	10	10	10	10	10	10	10	10	10	10	10
Two	Pearson Correlation	.999**	.963**	1	.998**	.998**	.998**	.998**	.998**	.998**	.998**	.998**
	Sig. (2-tailed)	.000	.000		.000	.000	.000	.000	.000	.000	.000	.000
	N	10	10	10	10	10	10	10	10	10	10	10
Three	Pearson Correlation	1.000**	.951**	.998**	1	1.000**	1.000**	1.000**	1.000**	1.000**	1.000**	1.000**
	Sig. (2-tailed)	.000	.000	.000		.000	.000	.000	.000	.000	.000	.000
	N	10	10	10	10	10	10	10	10	10	10	10
Four	Pearson Correlation	1.000**	.951**	.998**	1.000**	1	1.000**	1.000**	1.000**	1.000**	1.000**	1.000**
	Sig. (2-tailed)	.000	.000	.000	.000		.000	.000	.000	.000	.000	.000
	N	10	10	10	10	10	10	10	10	10	10	10
Five	Pearson Correlation	1.000**	.951**	.998**	1.000**	1.000**	1	1.000**	1.000**	1.000**	1.000**	1.000**
	Sig. (2-tailed)	.000	.000	.000	.000	.000		.000	.000	.000	.000	.000
	N	10	10	10	10	10	10	10	10	10	10	10
Six	Pearson Correlation	1.000**	.951**	.998**	1.000**	1.000**	1.000**	1	1.000**	1.000**	1.000**	1.000**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000		.000	.000	.000	.000
	N	10	10	10	10	10	10	10	10	10	10	10
Seven	Pearson Correlation	1.000**	.952**	.998**	1.000**	1.000**	1.000**	1.000**	1	1.000**	1.000**	1.000**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000		.000	.000	.000
	N	10	10	10	10	10	10	10	10	10	10	10
Eight	Pearson Correlation	1.000**	.951**	.998**	1.000**	1.000**	1.000**	1.000**	1.000**	1	1.000**	1.000**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000		.000	.000
	N	10	10	10	10	10	10	10	10	10	10	10
Nine	Pearson Correlation	1.000**	.952**	.998**	1.000**	1.000**	1.000**	1.000**	1.000**	1.000**	1	1.000**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000		.000
	N	10	10	10	10	10	10	10	10	10	10	10
Ten	Pearson Correlation	1.000**	.951**	.998**	1.000**	1.000**	1.000**	1.000**	1.000**	1.000**	1.000**	1
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	
	N	10	10	10	10	10	10	10	10	10	10	10

** Correlation is significant at the 0.01 level (2-tailed).

strings of data. In addition to the pages clicked by DMs, common to any study computing the CTRs of users, the algorithms provide a numerical representation of the evaluations determining their retrieval behavior. This information is not generally available, with most empirical studies aiming to extrapolate the preferences and evaluations of DMs from the data observed (Li et al., 2020; Wang et al., 2019). However, in the current setting, we obtain a numerical evaluation of each characteristic observed by the DMs, as well as the resulting actions in terms of clicking behavior.

The capacity of the benchmark algorithms to incorporate signals or any type of strategic behavior into the information retrieval process of DMs is novel to the current setting. That is, the algorithms allow to modify the beliefs or behavior of the DM at any exact point through the retrieval process and analyze the resulting consequences in terms of CTRs. The results presented in Table 4 provide intuition regarding the consequences derived from these strategic modifications throughout the search and evaluation process. Clearly, the difference in CTRs across retrieval scenarios is significant. Incorporating or eliminating a satisficing alternative shifts substantially the evaluation probabilities of those alternatives located in intermediate and lower ranking positions.

These results define two potential strategies that should be analyzed in future research. First, as stated in the introduction, the prospective benefits and costs derived from inducing relative shifts in the ranking position of the alternatives could be quantified and compared across a variety of retrieval scenarios. These changes would have a direct effect on the alternatives defining the corresponding node and all those located through the subsequent branches of the decision tree. Second, the preferences of DMs could be modified, inducing a change in their evaluation probabilities at any node composing the decision tree. The resulting probabilities are comparable across scenarios, providing firms whose products define the corresponding alternatives with the capacity to design marketing strategies depending on the scenario considered and the relative position of their products.

We conclude by noting that the decision model described in the paper does not consider the formation of preferences among DMs or the factors determining their choices. The strategies of firms to try biasing the choices of consumers towards their products have been consistently analyzed in the search engine marketing literature (Rosário & Dias, 2023). Future research should therefore analyze the strategic framework conditioning the retrieval behavior of DMs and their potential consumption choices.

In this regard, intuition can be drawn from the CTRs summarized in Tables 3 and 4. Consider the benchmark algorithm framework and the differences in CTRs described in both tables. If DMs evaluate a lower number of alternatives while exhibiting a higher rejection probability, they will tend to focus only on the four initial results displayed by the search engine.

This latter extension would lead the research into the domain of preference manipulation. Deciding what information to gather as well as the number of sources consulted based on the information already retrieved – formalized together with the limited capacity of DMs to assimilate information and search for alternative sources – delineates another potential line of research (Di Caprio & Santos-Arteaga, 2011). For instance, Epstein et al. (2017) illustrated how the search engine manipulation effect could be smoothed by issuing alerts to DMs, though it could only be eliminated by alternating the results of the search. They also concluded that this manipulation effect could impact multiple decision-making areas and suggested the regulation of search engines.

Current literature developments incorporate the effect of artificial intelligence on the information retrieval behavior of DMs (Verma et al., 2021). This recent research trend also analyzes the importance of corporate digital responsibility when shifting from Industry 4.0 to Industry 5.0 (Kraus et al., 2022; Pappas et al., 2023). The latter integrates digital technologies with the problem-solving capacity of DMs, enhancing their ability to deal with and solve complex problems (Vassilakopoulou et al., 2023). The algorithms introduced in the current paper can be easily incorporated into these frameworks of analysis, allowing also for fuzzy extensions in the design of retrieval strategies coupled with credibility considerations, both of which are particularly relevant when accounting for sustainable scenarios (Pappas & Woodside, 2021; Santos-Arteaga et al., 2023).

7. Conclusion

We have defined a series of benchmark heuristic algorithms that mimic the information retrieval behavior of DMs and compared the CTRs obtained through different evaluation scenarios determined by the alternatives displayed on the first page of results delivered by a search engine. We have also demonstrated the enhanced performance of these algorithms when compared to a direct implementation of basic retrieval techniques. A formal information retrieval model has been introduced to relate the cutoff values defining each evaluation step of the algorithms to the preferences of DMs. We have illustrated how both processes can be compatible, combining the main ideas from rational decision making theory and bounded rationality within a verifiable context of online search behavior.

The benchmark algorithms have been designed using simple behavioral constraints, allowing for the integration of complex strategies incorporating signals, subjective beliefs, and exogenous preference shocks to the information retrieval process of DMs. The benchmarking quality of the algorithms allows to simulate a substantial number of potential scenarios describing the effects from changes in the willingness of DMs to click on a given set of alternatives. As a result, we can quantify the expected adjustments in CTRs arising from different modifications introduced at any point through the information retrieval process of DMs. One of the main drawbacks inherent to the structure of the benchmark algorithms relates to the complexity that results from the inclusion of additional alternatives in the retrieval process of DMs.

We conclude by emphasizing that the algorithms are sufficiently malleable to consider the content of the websites linked together with the description provided by the engine as the features of the alternatives aligning with the preferences of DMs. In this regard, a second stage per node could be incorporated to the different decision-tree structures, doubling the number of nodes composing the algorithms and the subsequent retrieval probabilities. Note also that the resulting prescriptive environment can be used as a benchmark relative to which DMs can be categorized when training any of the deep learning techniques commonly applied in the artificial intelligence literature.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Dr. Francisco J. Santos-Arteaga is grateful for the support received from the María Zambrano contract from the Universidad Complutense de Madrid financed by the Ministerio de Universidades with funding from the European Union Next Generation program.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jjimei.2024.100229.

References

Advanced Web Ranking: Google organic CTR history. Last accessed December 23rd, 2021. (2021) https://www.advancedwebranking.com/ctrstudy/.
Baclawski, K. (2018). The observer effect. In *IEEE conference on Cognitive and Computational Aspects of Situation Management (CogSIMA)* (pp. 83–89). https://doi.org/10.1109/COGSIMA.2018.8423983
Baeza-Yates, R.: Applications of web query mining. In: Losada, D.E., Fernández-Luna, J. M. (Eds.). Advances in information retrieval. information systems and applications, incl. internet/web, and HCI, Vol. 3408. Springer-Verlag, Berlin Heidelberg, 7–22 (2005).
Basu, S. (2018). Information search in the internet markets: Experience versus search goods. *Electronic Commerce Research and Applications, 30*, 25–37.
Bossaerts, P., & Murawski, C. (2015). From behavioural economics to neuroeconomics to decision neuroscience: The ascent of biology in research on human decision making. *Current Opinion in Behavioral Sciences, 5*, 37–42.
Chen, S. Y., & Macredie, R. (2010). Web-based interaction: A review of three important human factors. *International Journal of Information Management, 30*(5), 379–387.
Chitika. (2013). The value of Google result positioning. *Chitika Insights*. June 7, 2013. Chitika, WestboroughAvailable at perma.cc/7AGC-HTDH.
Dean, B.: We analyzed 5 million Google search results. Here's what we learned about organic click through rate (2019) Available at https://backlinko.com/google-ctr-stats.
Di Caprio, D., & Santos-Arteaga, F. J. (2011). Strategic diffusion of information and preference manipulation. *International Journal of Strategic Decision Sciences, 2*(2), 1–19.
Di Caprio, D., & Santos-Arteaga, F. J. (2021). Combinatorial abilities and heuristic behavior in online search environments. *Operations Research Perspectives, 8*, Article 100179.
Di Caprio, D., Santos-Arteaga, F. J., & Tavana, M. (2022a). An information retrieval benchmarking model of satisficing and impatient users' behavior in online search environments. *Expert Systems with Applications, 191*, Article 116352.
Di Caprio, D., Santos-Arteaga, F. J., & Tavana, M. (2022b). A new algorithm for modeling online search behavior and studying ranking reliability variations. *Applied Intelligence, 52*, 7529–7549.
Epstein, R., & Robertson, R. E. (2015). The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. *Proceedings of the National Academy of Sciences, 112*, E4512–E4521.
Epstein, R., Robertson, R. E., Lazer, D., & Wilson, C. (2017). Suppressing the search engine manipulation effect (SEME). *Proceedings of the ACM on Human-Computer Interaction, 1*(CSCW), 1–22.
European Commission. (2016). *Online platforms. special eurobarometer 447*. European Union. April 2016.
Gönsch, J. (2020). How much to tell your customer? – A survey of three perspectives on selling strategies with incompletely specified products. *European Journal of Operational Research, 280*(3), 793–817.
Guan, Z., & Cutrell, E. (2007). An eye tracking study of the effect of target rank on web search. In *Proceedings of the special interest group for computer-human interaction conference on human factors in computing systems* (pp. 417–420). New York: ACM.
Ham, J., Lee, K., Kim, T., & Koo, C. (2019). Subjective perception patterns of online reviews: A comparison of utilitarian and hedonic values. *Information Processing & Management, 56*(4), 1439–1456.
Hendahewa, C., & Shah, C. (2017). Evaluating user search trails in exploratory search tasks. *Information Processing & Management, 53*, 905–922.
Herrmann, J. W. (2015). *Engineering decision making and risk management*. Hoboken, New Jersey: John Wiley & Sons, Inc.
Hong, W., Xiong, Z., You, J., Wu, X., & Xia, M. (2021). CPIN: Comprehensive present-interest network for CTR prediction. *Expert Systems with Applications, 168*, Article 114469.
Jain, S., Seeja, K. R., & Jindal, R. (2021). A fuzzy ontology framework in information retrieval using semantic query expansion. *International Journal of Information Management Data Insights, 1*(1), Article 100009.
Jansen, M. B. J., Spink, A., Bateman, J., & Saracevic, T. (1998). Real life information retrieval: A study of user queries on the web. *ACM SIGIR Forum, 32*, 5–17.
Joseph, J., & Gaba, V. (2020). Organizational structure, information processing, and decision-making: A retrospective and road map for research. *Annals, 14*, 267–302.
Kabiawu, O., van Belle, J.-P., & Adeyeye, M. (2016). Designing a knowledge resource to address bounded rationality and satisficing for ICT decisions in small organizations. *The Electronic Journal of Information Systems in Developing Countries, 73*, 1–18.
Kim, Y. J., Song, J., & Koo, C. (2008). Exploring the effect of strategic positioning on firm performance in the e-business context. *International Journal of Information Management, 28*(3), 203–214.
Kraus, S., Durst, S., Ferreira, J. J., Veiga, P., Kailer, N., & Weinmann, A. (2022). Digital transformation in business and management research: An overview of the current status quo. *International Journal of Information Management, 63*, Article 102466.
Lewandowski, D., & Kammerer, Y. (2021). Factors influencing viewing behaviour on search engine results pages: A review of eye-tracking research. *Behaviour & Information Technology, 40*(14), 1485–1515.
Li, H., Duan, H., Zheng, Y., Wang, Q., & Wang, Y. (2020). A CTR prediction model based on user interest via attention mechanism. *Applied Intelligence, 50*, 1192–1203.
Lim, S. (2013). College students' credibility judgments and heuristics concerning Wikipedia. *Information Processing & Management, 49*, 405–419.
List, A., & Alexander, P. A. (2017). Text navigation in multiple source use. *Computers in Human Behavior, 75*, 364–375.

Lorigo, L., Haridasan, M., Brynjarsdóttir, H., Xia, L., Joachims, T., Gay, G., et al. (2008). Eye tracking and online search: Lessons learned and challenges ahead. *Journal of the American Society for Information Science and Technology, 59*, 1041–1052.

Luo, W., Cook, D., & Karson, E. J. (2011). Search advertising placement strategy: Exploring the efficacy of the conventional wisdom. *Information & Management, 48*, 404–411.

Miranda, A., & Miah, S. J. (2023). Designing an innovative unified contextual architecture for improving information retrieval service in healthcare organizations. *Information Development, 39*(2), 235–247.

Mishra, R. K., Jothi, J. A. A., Urolagin, S., & Irani, K. (2023). Knowledge based topic retrieval for recommendations and tourism promotions. *International Journal of Information Management Data Insights, 3*(1), Article 100145.

O'Brien, H. L., Arguello, J., & Capra, R. (2020). An empirical study of interest, task complexity, and search behaviour on user engagement. *Information Processing & Management, 57*(3), Article 102226. https://doi.org/10.1016/j.ipm.2020.102226

Pan, B., Hembrooke, H., Joachims, T., Lorigo, L., Gay, G., & Granka, L. (2007). In Google we trust: Users' decisions on rank, position, and relevance. *Journal of Computer-Mediated Communications,, 12*, 801–823.

Pappas, I. O., Mikalef, P., Dwivedi, Y. K., Jaccheri, L., & Krogstie, J. (2023). Responsible digital transformation for a sustainable society. *Information Systems Frontiers, 25*, 945–953.

Pappas, I. O., & Woodside, A. G. (2021). Fuzzy-set qualitative comparative analysis (fsQCA): Guidelines for research practice in information systems and marketing. *International Journal of Information Management, 58*, Article 102310.

Qin, J., Zhang, W., Wu, X., Jin, J., Fang, Y., & Yu, Y. (2020). User behavior retrieval for Click-Through Rate prediction. In *Proceedings of the 43rd International ACM SIGIR conference on research and development in information retrieval (SIGIR '20)* (pp. 2347–2356). New York, NY, USA: Association for Computing Machinery.

Ren, H., & Huang, T. (2018). Modeling customer bounded rationality in operations management: A review and research opportunities. *Computers & Operations Research, 91*, 48–58.

Rosário, A. T., & Dias, J. C. (2023). How has data-driven marketing evolved: Challenges and opportunities with emerging technologies. *International Journal of Information Management Data Insights, 3*(2), Article 100203.

Santos-Arteaga, F. J., Di Caprio, D., Tavana, M., & Cerda Tena, E. (2023). A credibility and strategic behavior approach in hesitant multiple criteria decision-making with application to sustainable transportation. *IEEE Transactions on Fuzzy Systems, 31*(2), 460–474.

Schmitt, J. B., Debbelt, C. A., & Schneider, F. M. (2018). Too much information? Predictors of information overload in the context of online news exposure. *Information, Communication & Society, 21*(8), 1151–1167.

Schulz, F. (2008). Trees with exponentially growing costs. *Information and Computation, 206*, 569–578.

Schwartz, B. (2005). *The paradox of choice: Why more is less*. Harper Perennial.

Sharma, S., Rana, V., & Kumar, V. (2021). Deep learning based semantic personalized recommendation system. *International Journal of Information Management Data Insights, 1*(2), Article 100028.

Shepherd, S. J. (2007). Concepts and architectures for next-generation information search engines. *International Journal of Information Management, 27*(1), 3–8.

Silverstein, C., Marais, H., Henzinger, M., & Moricz, A. (1999). Analysis of a very large web search engine query log. *Association for Computing Machinery Special Interest Group on Information Retrieval Forum, 33*, 6–12.

Speier-Pero, C. (2019). Using aggregated data under time pressure: A mechanism for coping with information overload. *Journal of Decision Systems, 28*, 82–100.

Sun, H., Fan, M., & Tan, Y. (2020). An empirical analysis of seller advertising strategies in an online marketplace. *Information Systems Research, 31*, 37–56.

Sutanto, J., Liu, Y., Grigore, M., & Lemmik, R. (2018). Does knowledge retrieval improves work efficiency? An investigation under multiple systems use. *International Journal of Information Management, 40*, 42–53.

Tavana, M., Di Caprio, D., & Santos Arteaga, F. J. (2016a). Modeling sequential information acquisition behavior in rational decision making. *Decision Sciences, 47*(4), 720–761.

Tavana, M., Santos-Arteaga, F. J., Di Caprio, D., & Tierney, K. (2016b). Modeling signal-based decisions in online search environments: A non-recursive forward-looking approach. *Information & Management, 53*(2), 207–226.

Tavana, M., Di Caprio, D., & Santos Arteaga, F. J. (2015a). An ordinal ranking criterion for the subjective evaluation of alternatives and exchange reliability. *Information Sciences, 317*, 295–314.

Tavana, M., Di Caprio, D., Santos Arteaga, F. J., & O'Connor, A. (2015b). A novel entropy-based decision support framework for uncertainty resolution in the initial subjective evaluations of experts: The NATO enlargement problem. *Decision Support Systems, 74*, 135–149.

Utku Özmen, M., & Yucel, E. (2019). Handling of online information by users: Evidence from TED talks. *Behaviour & Information Technology, 38*(12), 1309–1323.

Vassilakopoulou, P., Haug, A., Salvesen, L. M., & Pappas, I. O. (2023). Developing Human/AI interactions for chat-based-customer-services: Lessons learned from the norwegian government. *European Journal of Information Systems, 32*(1), 10–22.

Verma, S., Sharma, R., Deb, S., & Maitra, D. (2021). Artificial intelligence in marketing: Systematic review and future research direction. *International Journal of Information Management Data Insights, 1*(1), Article 100002.

Victorelli, E. Z., Dos Reis, J. C., Hornung, H., & Prado, A. B. (2020). Understanding human-data interaction: Literature review and recommendations for design. *International Journal of Human-Computer Studies, 134*, 13–32.

Walgrave, S., & Dejaeghere, Y. (2017). Surviving information overload: How elite politicians select information. *Governance, 30*, 229–244.

Wang, Q., Liu, F., Xing, S., & Zhao, X. (2019). Research on CTR prediction based on stacked autoencoder. *Applied Intelligence, 49*, 2970–2981.

Wu, I.-C., & Yu, H.-K. (2020). Sequential analysis and clustering to investigate users' online shopping behaviors based on need-states. *Information Processing & Management, 57*(6), Article 102323. https://doi.org/10.1016/j.ipm.2020.102323

Yang, L., Toubia, O., & De Jong, M. G. (2015). A bounded rationality model of information search and choice in preference measurement. *Journal of Marketing Research, 52*(2), 166–183.

Yu, H., Jatowt, A., Blanco, R., Joho, H., & Jose, J. M. (2017). Decoding multi-click search behavior based on marginal utility. *Information Retrieval Journal, 20*, 25–52.

Zanganeh, M. Y., & Hariri, N. (2018). The role of emotional aspects in the information retrieval from the web. *Online Information Review, 42*(4), 520–534.

Zhang, J., Adomavicius, G., Gupta, A., & Ketter, W. (2020). Consumption and performance: Understanding longitudinal dynamics of recommender systems via an agent-based simulation framework. *Information Systems Research, 31*, 76–101.