No Evidence of Attraction Effect Among Recommended Options: A large-scale field experiment on an online flight aggregator[+]

Ismael Rafai (Université Côte d'Azur)[*] Zakaria Babutsidze (Université Côte d'Azur)

Thierry Delahaye (Amadeus) Nobuyuki Hanaki (Osaka University)

Rodrigo Acuna-Agost (Amadeus)

September 20, 2021

Abstract. This article is the first to test whether the attraction effect can be replicated among recommended options in a real digital marketplace, namely – an online flight aggregator. For this purpose, we conducted a large-scale field experiment on an existing flight aggregator, in which we varied the number of the options recommended at the top of the result list. More precisely, we investigated whether recommending an additional "decoy" option, which is asymmetrically dominated by one of the two former recommended options (the cheapest or the fastest itinerary), increases users' conversion rate and impacts the market share of the recommended options. Our results, based on the analysis of more than 140,000 search sessions, suggest that this is not the case in general. We then conduct analyses on subsamples to investigate the boundary conditions of the attraction effect. This study questions the relevance of the attraction effect in online marketplaces and recommender systems and proposes new research avenues.

Keywords. *Digital nudge, E-commerce, Recommender system, Flight booking, Behavioral bias, Attraction Effect*

[+] **Acknowledgment.** This research was supported by the French National Research Agency (ANR) (GRANT: ANR-15-IDEX-01). Data for this research was provided by Amadeus, a major IT provider for the global travel and tourism industry. We thank the anonymous reviewers, for their useful comments that helped improve this work. We thank Catherine Buchanan for proofreading this article.
[*] corresponding author: ismael.rafai@gmail.com

Introduction

With the proliferation of the Internet, purchases are increasingly made through online marketplaces. These outlets offer consumers a wider variety of products than most brick-and-mortar stores can afford. However, this increase in product range is only beneficial to consumers if they can identify the products they prefer from among the set available. To address this issue, a large body of the information systems literature has been dedicated to the study and improvement of recommender systems (Bobadilla et al., 2013; Lu et al., 2015). Recommender systems are "programs which attempt to recommend the most suitable *items* (products or services) to particular users (individuals or businesses) by predicting a user's interest in an item based on related information about the items, the users and the interactions between items and users" (Lu et al., 2015). Recommender systems, thus, aim to assist the users of digital platforms, by identifying and highlighting the alternatives that are most likely to match users' needs.

If consumers were utility maximizers, as assumed by the standard rational choice theory, this matching would be assured by identifying the distribution of users' preferences in order to recommend the products that are most likely to be preferred. However, there is plenty of evidence from economics, psychology and management science arguing that humans do not behave as simple preference maximizers, but use various heuristics (Babutsidze, 2012) and are subject to decision biases (Tversky & Kahneman, 1975). These departures from rationality reflect the fact that preferences, or at least their expression through choices, are not neutral with respect to the choice set presentation. Applied to information systems research, the interaction between preferences and the choice presentation should be considered when designing a recommender system (Adomavicius et al., 2019): this interaction could be either limited – to avoid negative side-effects – or on the contrary enhanced, to improve recommender system performance.

Since choices are impacted by the way they are presented, the choice architect – i.e., the designer of the choice environment – may *nudge* users. By this we mean that the choice designer is able to change "the aspect of the choice architecture [to] alter people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives" (Thaler & Sunstein, 2009). This concept of nudges, coming from psychology and behavioral economics, has been recently introduced and is becoming popular in information systems research (David Arnott & Gao, 2019; Jameson et al., 2014; Weinmann et al., 2016).

Despite digital nudging gaining popularity among practitioners, the vast majority of empirical studies to date were conducted in laboratory settings. Laboratory studies are ideal to prove the existence and assess the internal validity of behavioral biases, and to identify variables and conditions that moderate their effects. However, once internal validity of the bias is assessed, and once it has been replicated several times in the laboratory, it is also important to test if the bias is relevant in the field. Indeed, the relevance of a behavioral bias in the (digital) field cannot simply be assumed as demonstrated by the results presented in this article.

The digital nudge we are interested in, has principally been tested in situations where the researcher can design an ideal environment where all alternatives' attributes are defined and controlled to maximize the likelihood of observing the effect. On the contrary, in the field, many recommender systems cannot control product and service attributes, but rather can only control how they are disclosed/presented. Moreover, in the laboratory, decision makers (i) know they are participating in an experiment, (ii) choose from a very limited number of alternatives, with limited consequences, and (iii) are "captive", in the sense that they do not have many outside options (including the possibility to delay their choices or to choose from another, competing, platform). It is, therefore, fundamental to test the conceptual replicability of lab-identified behavioral biases in the field in order to assess their practical relevance (Hüffmeier et al., 2016).

In this article, we present the first test of the effectiveness of a "digital decoy nudge" in a real recommender system setting. We rely on a large-scale field experiment and observe more than 140,000 searches made on an online flight aggregator, where we are able to manipulate the set of recommended itineraries. The digital decoy nudge we are interested in is based on a decision bias known as the asymmetric dominance effect, or attraction effect (hereafter AE). It consists of recommending one "decoy" itinerary that is dominated by one of the former recommended itineraries (the target), but not by others, in order to increase market share of the target recommendation. Our paper, therefore, contributes both to the literature on digital nudges, and to the literature on the AE. In addition, while many behavioral biases are conceptually easy to understand, operationalizing and testing them in the digital field represents a significant organizational and cultural challenge from a company point of view (Gupta et al., 2019). As a result, this article may interest companies aiming to implement digital nudges in their recommender systems, by providing an example of how a digital decoy nudge could be operationalized into a search engine.

In our results, we do not detect the presence of the attraction effect on recommended options in the specific setting studied. However, our large dataset allows us to control for session characteristics and (at least partially) rule out potential explanations for the absence of the effect. For example, we found that manipulating the size of the choice set does not affect the size of the attraction effect, nor does the similarity between the target and the competitor options, the stakes of the decision, or the device used (computer vs smartphone) in the search process. While questioning the practical relevance of the effect for recommender systems, additional exploratory results suggest that it could be possible to identify the source of effect heterogeneity in future research, paving the way to the implementation of personalized digital nudges in recommender systems.

The rest of the paper is organized as follows. The next section reviews existing literature on digital nudges and on the AE, discusses the relevance of the effect for recommender systems, and presents our hypotheses. We then present the details of our experiment and its results. The last section concludes with a discussion and a presentation of potential avenues for future research.

Literature review

Digital nudges

In theory, the digital world is particularly suited for designing, testing and evaluating the impact of nudges, since the choice architect can easily manipulate "user-interface design elements to guide people's behavior in digital choice environments" (Weinmann et al., 2016) and can observe nearly immediately the impact of the digital nudge on behaviors.[1] There are many examples of the use of digital nudges. For example, in e-commerce, Adomavicius et al., (2013) report the presence of anchoring effects; Camilleri (2020) highlights an order effect between reviews and product presentation on purchase intentions; Shen et al., (2020) show that

[1] Originally, in behavioral economics, nudges designated interventions that were intended to improve individuals' welfare, and the term "sludge" was preferred when similar techniques were used to increase a seller's profit by negatively impacting consumers. However, in the information system literature, the term "digital nudge" has been introduced to designate small modifications of the choice environment that influence users' decisions, regardless of their impact on welfare (Schneider & Weinmann, n.d.).

introducing a distraction task between products' presentation may facilitate complex product choices by evoking unconscious thoughts; and Sheng & Joginapelly (2012) show that purchase intentions can be enhanced by the presence of web atmospheric cues on an e-commerce website. Similarly, Liu & Karahanna (2017) show that online reviews influence the weight of attributes in product evaluation; Aerts et al. (2017) show that website design may nudge reviewers to deliver higher quality reviews and Craciun et al. (2020) show that the perception of a review is impacted by the presentation of the reviewer's gender. Schneider et al. (2021) report that the overall rating of a product can be manipulated by proposing multi-dimensional rating systems, where consumers are also asked to rate precise attributes of the product.

In reward-based crowdfunding, Wessel et al. (2018) demonstrate that unavailable sold-out early bird options impact option choices. Adomavicius et al. (2018) show that the willingness to pay for digital songs is impacted by systems' recommendations. In the context of mobile phone recommendations, Ho & Lim (2018) found that hiding products' attributes increases consumers' urge to buy and thus their unplanned purchases, and that this effect interacts with consumers' mood. Kroll & Stieglitz (2019) show that digital nudging can be used to increase self-disclosure in social networks. Sivan et al. (2019) report that the position of legal alternatives in a recommender system reduces piracy behavior.[2]

Attraction Effect

A particular digital nudge on which we focus in this paper is based on a decision bias known as the asymmetric dominance effect, or attraction effect. The AE was introduced by Huber et al. (1982) and predicts the increase of the choice probability of one alternative (the target) caused by the introduction of a new alternative (the decoy) in the choice set. For this to happen, the decoy should be dominated by the target but not by competing alternatives (that is, the decoy is inferior in at least one characteristic and is not superior in any characteristic to the target; but the decoy is superior to other competing alternatives in at least one characteristic), see Figure 1 for an example. The AE has received great attention from both professionals (because of its marketing possibilities) and academics (as it constitutes a major departure from the standard

[2] Other examples can be found in Arnott & Gao (2019).

rational choice theory).[3]

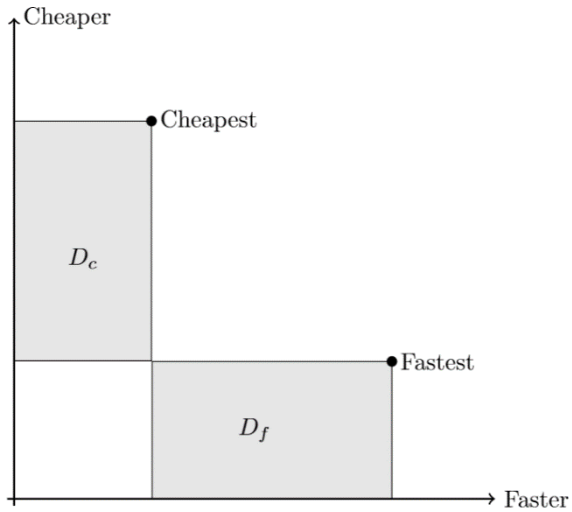


Figure 1: Example of potential decoys' positions for itinerary choices.

Note: In this itinerary choice example, alternatives are defined by only two attributes: price and duration (the cheaper and the faster, the better). Alternatives located in the D_c region are dominated by the Cheapest itinerary (they are more expensive but not faster than the Cheapest), but not by the Fastest itinerary (they are slower but cheaper than the Fastest) and are potential decoys targeting the Cheapest itinerary. Symmetrically, the alternatives located in the D_f region are potential decoys targeting the Fastest.

Explanations and moderators of the attraction effect

The AE has been widely documented in the laboratory, as well as in survey settings, including purchasing goods (Doyle et al., 1999a), lottery choices (Castillo, 2020), perceptual tasks (Trueblood et al., 2013), and even animal decision making (Shafir et al., 2002).[4] Several explanations of the effect have been proposed, which may be grouped into compensatory and non-compensatory decision strategies. Compensatory strategies assume that individual decisions are based on a trade-off between the product attributes (Lee & Anderson, 2009). Within this framework, the AE can be explained by a change in the weighting of the attributes, induced by the introduction of the decoy. Indeed, the introduction of the decoy can either (i) extend the range

[3] From a theoretical point of view, the AE implies a violation of the Regularity condition (the probability of choosing an alternative cannot increase when the choice set is extended), which is incompatible with any random utility model (Dasgupta & Pattanaik, 2007).

[4] See Lichters et al. (2015) for a review.

of attribute's values, narrowing the perceived distance between options within the range (range theory) or (ii) increase the number of levels available to define the attribute, increasing its weight in the decision (frequency theory). However, explanations based on compensatory choices cannot explain why individual differences moderate the effect, and in particular, why the AE is stronger for younger individuals, those with lower levels of literacy, cognitive ability, and higher faith in intuitions (Kim & Hasher, 2005; Mao & Oppewal, 2012). Neither are they able to explain why the AE does not systematically increase with increasing range of the decoy attribute (Castillo, 2020).

In contrast, other explanations of the AE are based on a modification of the decision process: the introduction of a decoy, and the asymmetric dominance relation it creates, offers individuals new ways to make and justify their decisions. These explanations can be related to non-compensatory strategies. Non-compensatory strategies are based on heuristics rather than attribute trade-offs. For example, a consumer may decide to restrict the choice only to the N alternatives with the lowest price, or with the highest quality, or to the options that are unambiguously dominating other options. In those conditions the introduction of the decoy may imply that a competing alternative, which would have been preferred to the target if the consumer was using a compensatory strategy, would be excluded from the consideration set altogether. Overall, while there is no single (type of) explanation that can encompass all moderators of AE that have been identified (Yang, 2013), the AE suggests that the decision process does not simply consist in revealing existing preferences. In particular, having strong priors, and the difficulty of perceiving the dominance relationship are two factors that crucially limit the AE.[5]

Ecological validity of the effect

Recently, some failures to replicate the AE, and the restrictive conditions under which the AE is usually obtained have raised a debate on the ecological validity of these studies. This questions the relevance and economic significance of the AE (Frederick et al., 2014; Huber et al., 2014; Yang & Lynn, 2014). To the

[5] Experimental manipulation and product presentation can moderate the attraction effect: the relevance and meaningfulness of the product decreases the AE (Mishra et al., 1993), while the use of pictorial and tactile forms (Simonson & Tversky, 1992), video demonstrations (Slaughter et al., 1999) or graphical representations (Choplin & Hummel, 2002) enhance the AE.

best of our knowledge, there is only one article that directly operationalizes the attraction effect in the field, with real customers (Doyle et al., 1999a). The authors investigated the AE in a local grocery store, by creating dominated baked beans (either by introducing a more expensive brand of the same quality as the higher quality baked beans, or low-quality beans with damaged cans into the choice set) and observing purchases from real customers.[6] Other studies investigate the AE by using a hypothetical two-alternative (plus a decoy) forced-choice paradigm, where alternatives are defined through two numerical attributes (e.g., price and quality) upon which consumers' preferences are straightforwardly defined (almost all consumers prefer low price and high quality). In such a design, real incentives are absent (Lichters et al., 2017), the size of the choice set is limited (Dimara et al., 2017), more subjective attributes are absent, and it is impossible to abstain from choice (Dhar & Simonson, 2003). These are all factors that are not applicable to most marketplaces but have been suspected to be necessary for consumers to identify dominance and to act on it (Bettman et al., 1998). Following this debate, Lichters et al. (2015) have proposed guidelines to conduct experiments on the AE to increase both the likelihood of obtaining the effect, and the external validity of the studies. These guidelines include the presence of economic consequences, a realistic and meaningful product presentation, presence of a no-buy option, relevance of the sample, and the avoidance of artificially repeated choices.

Testing the AE in a real recommender system

While Huber et al. (2014) admitted that computerized laboratory studies can lack external validity to explain purchases made in brick-and-mortar stores, where customers can physically interact with the products they want to purchase, they argued that the criticism is less true for digital marketplaces. In digital marketplaces, product descriptions are closer to how choices are presented in computerized experiments, where product quality is often reduced to a numerical rating. Therefore, the emergence of digital marketplaces should increase the likelihood of situations occurring that are favorable to the appearance of the AE. In this desire to increase external validity, some papers proposed to test the AE with experiments that closely follow the guidelines by Lichters et al. (2015). These experiments are referred to as "artificial field experiments" (online experiments

[6] Another study conducted by Li et al. (2019) proposed to test the attraction effect in three field experiments to increase hand hygiene in a food factory, but has been retracted due to anomalies and untraceable data collection processes.

conducted on a more representative sample than students) and “framed field experiments” (online experiments that mimic or replicate the actual design of existing digital marketplaces). For example, Weinmann et al. (2020) highlight the AE in reward-based crowdfunding by running experiments on Prolific and by replicating a Kickstarter campaign, with economic consequences. Gomez et al. (2016) mimic websites like Phillips.com and Mediamarkt.com for choices among portable speakers and Groupon.com for choices among restaurant menus. Those experiments strongly suggest that the AE could appear in the digital field.

However, despite their high level of external validity, framed field experiments cannot reach the same level as real field or natural experiments. The respondents of framed field experiments are volunteers aware of their participation in an experiment. Therefore, they can be subject to experimenter demand effects (Zizzo, 2010). Even if the choice is made real by the presence of economic incentives, and the participant is not forced to buy thanks to the presence of a no buy option, the context of the choice remains artificial in the sense that the participants are not real customers interested in buying the product outside of the study, and they may use different decision processes than the ones they would use in their everyday lives.

In addition, there are many features of online marketplaces that are absent from both offline marketplaces and framed field experiments. In online marketplaces, consumers are less captive and are free to search on other websites, implying that the set of outside options is much wider than the one proposed in the existing studies (given that only few studies actually proposed an outside option (Dhar & Simonson, 2003; Lichters et al., 2017)). Another distinguishing feature is the dramatically high number of alternatives that users of online marketplaces have to search through. For instance, the only previous study testing AE in the field was conducted in a local grocery store (Doyle et al., 1999b) and was concerned with three types of baked beans (one target, one competitor and one decoy) that were available in the store. Importantly, in such a setting, consumers can be considered as (more) captive, very differently from digital marketplaces -- it is costlier to move from a store to another just because the choice set of baked beans is not fully satisfying to a consumer, than moving from a website to another.

Moreover, the AE has always been investigated in situations where the choice architect can freely define or control product attributes. One strength of such laboratory experiments is that the researcher can introduce the decoy that is expected to maximize the attraction effect (Kaptein et al., 2016). In contrast, many recommender systems cannot directly change the choice set by creating new options or redefining their attributes, but acts only on the saliency of the existing options (e.g., by highlighting or sorting those options on top) to increase their likelihood of being considered. Since being considered is a necessary step for an option to be chosen, highlighting existing decoys (that would have not been perceived previously) should have qualitatively similar effects than introducing new decoys in the choice set (Brady & Rehbeck, 2016). In particular, the existence of "phantom decoy effects" (the decoy alternative is presented but unavailable) proves that the availability of the decoy is not a necessary condition for the AE to be observed (Pettibone & Wedell, 2007). However, one may expect the effect to be reduced if compared to the standard decoy experiment because (i) some of the users may consider the decoy even when it is not recommended, and (ii) the attributes of the recommended decoy cannot be defined to maximize the effect size. On the other hand, in the situations where many alternatives are available, which are the ones where recommender systems are the most useful, consumers are suspected to adopt more non-compensatory choice processes, and heuristics, and the AE is thus expected to increase (Banerjee et al., 2020).

Finally, most of the studies investigating the AE focused on the impact of the decoy on the probability of choosing the target, relative to a specific competitor. This makes sense for sellers who are only interested in their own sales. However, from the point of view of a recommender system, it is interesting to study the impact of recommending a decoy, not only on the probability of choosing the target, but also on the probability of choosing the other recommended or available options. Therefore, in this paper, we will test the following three hypotheses:

H1: recommending a decoy option increases the probability that the user will choose one of the available options out of all options available on the platform.

H2: recommending a decoy option increases the probability that the user will choose one of the recommended options.

H3: recommending a decoy option increases the probability that the user will choose the targeted option (i.e., a recommended option that dominates the decoy).

The study

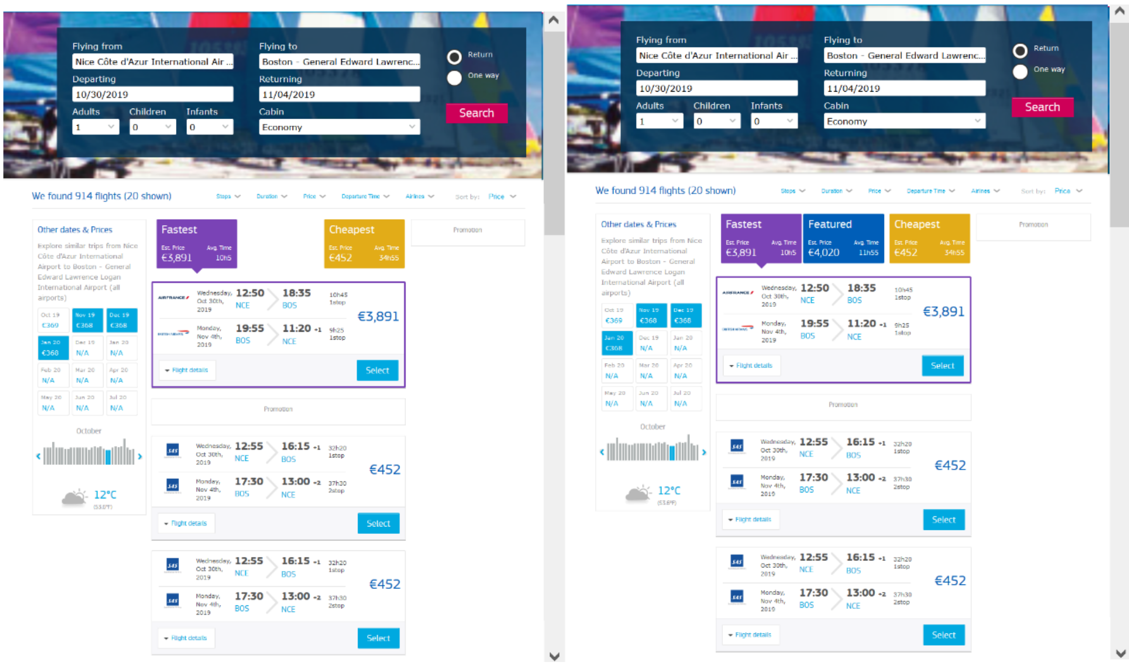
Our study aims to investigate the external validity of the AE on recommended options in a real online marketplace: namely online flight booking. More precisely, we focus here on a consumer's decision to book (or not) one of the available alternatives, from a menu comprising all flight itineraries from airport *A* to airport *B* on date *T* (and potentially a return on date *T'*) proposed and ranked by a search engine. As discussed above, such an environment contrasts with the standardized paradigm that has been previously used to study the AE. The consumer faces a large range of alternatives, which vary in numerous vertical (e.g., price and duration) and horizontal (e.g., airline and time of departure) attributes, and is not forced to make a choice. If we set aside consumers' preferences among the horizontal dimensions and focus on price and duration, consumers typically face a time-money trade-off whenever the cheapest available itinerary is not the fastest one. Usually, those two options are recommended to help consumers to make their choice and we can thus investigate whether the likelihood of choosing a target itinerary – e.g., the fastest or the cheapest option on the list – is increased when a decoy itinerary (i.e., an itinerary asymmetrically dominated in price or duration by the target itinerary but not by its competitors) is also recommended.

Field experiment on an online flight aggregator

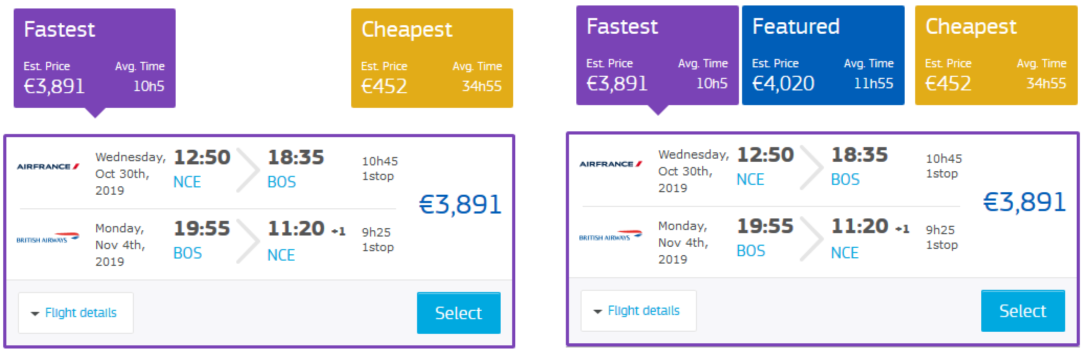
We conducted a large-scale field experiment in 2019 (from April 14th to May 24th and from July 31st to November 1st), on a real online flight aggregator. As in other flight aggregators, users indicate a departure date *D*, a departure airport *A* and a destination airport *B* (and potentially return date *D'*) to receive a list of all (relevant) available itineraries, offered by most online travel agencies and airlines worldwide. Above this list, a "highlighted panel" is displayed to recommend two itineraries - the "Cheapest" and the "Fastest".[7] Users are free to choose several itineraries from the list, or from the highlighted panel by clicking on the "Select" Button (Figure 2, left). If they do so, more information and a list of sellers (airline and/or travel agency) is provided, and the user can choose one of those sellers and be redirected to its website to perform a booking. We do not observe in our experiment the final booking decision, but the redirection to the sellers' website is considered

[7] In cases where the cheapest (respectively the fastest) itinerary was not unique, the recommender system highlighted the fastest of the cheapest itineraries (respectively the cheapest of the fastest itineraries).

as a “conversion” from the perspective of the aggregator. All in all, we observe a total of 200 992 search sessions, but focus the main analysis on the 146 851 sessions where at least one potential decoy itinerary was available (the analysis of the sessions where no potential decoy was available is presented in Appendix A). The next subsection explains how we identify the decoy and our experimental manipulation.



a) Screenshot of the baseline version. b) Screenshot of the decoy version.



c) Highlighted panel of the baseline version. d) Highlighted panel of the decoy version.

Figure 2: Screenshots of a search on the website.

Note: Left (right) figures show the NoDecoy (Decoy - here targeting the Fastest) versions of the website. More information about an itinerary is disclosed when clicking on the “Flight details” button. The “Select” button opens a list-box with several links to sellers’ websites for booking. The exact design depends on users’ browser and graphics settings. Users have to scroll down to see more alternatives and may apply a filter to mask or change the ranking of the itineraries. This does not affect the appearance of the highlighted itineraries.

Decoy identification and experimental manipulation

Our experimental manipulation consists in recommending a third "Featured" itinerary in the highlighted panel, as shown in Figure 2, which compares the "NoDecoy" setup, with two highlighted options, to the "Decoy" setup, with three highlighted ones.

This Featured itinerary is selected from the list of available options to be a decoy that targets either the Fastest (i.e., it is an itinerary more expensive than the Fastest, but faster than the Cheapest) or the Cheapest option (i.e., it is slower than the Cheapest, but cheaper than the Fastest), with the following procedure.

a) *Selection of the potential decoys targeting the Cheapest and the Fastest.*

Using the two recommended itineraries (the Cheapest and the Fastest), we identified two sets of potential decoy itineraries: (i) the set of decoy itineraries DeC (with price P_{DeC} and duration D_{DeC}) targeting the Cheapest, and (ii) the set of decoy itineraries DeF (with price P_{DeF} and duration D_{DeF}) targeting the Fastest. The itineraries in the two sets satisfy the following conditions:

- $P_C \leq P_{DeC} < P_F$ & $D_F < D_C \leq D_{DeC}$ for the set of decoys targeting the Cheapest, and
- $P_C < P_F \leq P_{DeF}$ & $D_F \leq D_{DeF} < D_C$ for the set of decoys targeting the Fastest,

where (P_C, D_C) and (P_F, D_F) denote the price and duration of the Cheapest and Fastest itineraries, respectively.

In the case where more than one decoy was targeting the Fastest (Cheapest), we select as a potential decoy targeting the Fastest (Cheapest) the decoy which was closest in departure time to the Fastest (Cheapest). Indeed, because departure time appeared to be one of the most important horizontal characteristics, we aimed to control for this dimension as much as possible.

b) *Selection of the Featured alternative.*

- Case 1: no potential decoy is targeting the Cheapest and no potential decoy is targeting the Fastest. In this case, the "Featured" itinerary was randomly selected from the set of all available itineraries. Those search sessions were different from the sessions where decoys were available and thus were removed from the main analysis. We analyze those search sessions separately in the appendix to assess the impact recommending a third alternative had on the conversion rates and found no impact of recommending a third alternative.

- Case 2: There is one potential decoy targeting the Cheapest OR one potential decoy targeting the Fastest, but not both. In this case, we choose this potential decoy as the Featured itinerary.
- Case 3: at least one potential decoy is targeting the Cheapest AND at least one potential decoy is targeting the Fastest. In this case, we randomly choose the Featured itinerary between the potential decoy targeting the Fastest and the potential decoy targeting the Cheapest with probability one half.

We identified a Featured itinerary for all search sessions, but it was highlighted only in the Decoy version of the website. We examine the impact of this experimental manipulation on three distinct conversion rates:

- the "overall conversion rate" (CR), which measures the probability that a user selects at least one of the available itineraries;
- the "highlighted conversion rate" (HCR), which measures the probability that a user selects at least one itinerary among the Faster or the Cheapest;
- the "targeted conversion rate", which measures the probability that a user selects the itinerary targeted by the decoy.

A statistically significant increase in those conversion rates, when recommending a decoy, would be interpreted as an evidence of the AE in online flight booking.

Control for session characteristics

The heterogeneity of search sessions allows us to study in which context the AE is the strongest. For example, Gaudeul & Crosetto (2019) show that the AE is more likely to occur when the competitor and the target alternatives are close in terms of utility and when the decisions are made quickly decisions, suggesting the AE is due to an intuitive cognitive process. Therefore, we can investigate if the AE is more likely to occur when the competitor and the target itineraries are similar or dissimilar, and when the travel is being planned in advance or not. We also investigate the impact of the size of the results list on the AE, since having more choice may either prevent the identification of the dominance relation (reducing the AE) or may incite the user toward more intuitive thinking (increasing the AE). Since we can identify if the users made their research using a mobile (tablet or phone) or a computer, we can also test if the AE is dependent on the device used. In particular, we collected and studied the impact of the following characteristics: (i) the “device used”: if the research was made on a *computer* or on a *mobile*; (ii) if the user was looking for a *one-way* or a return flight; (iii) the duration of the stay (only relevant for return travel), that is the difference between the departure date and the date of return; (iv) the number of days before the departures (i.e., the difference between the departure date and the date of the search); (v) the *menu size*, that is the total number of itineraries presented by the flight aggregator; (vi) the duration and the price of the cheapest and the fastest itineraries. Table 1 presents the characteristics of the sessions and shows that those characteristics are not statistically significantly different between the two versions of the website.

Variable	Definition	NoDecoy version		Decoy version		p-values (z-test)
		M	SD	M	SD	
Mobile	=1 If the search was made from a mobile device	27.26%	0.4453	27.56%	0.4468	0.4705
One way	=1 If the search did not include return itinerary	28.84%	0.4530	28.92%	0.4534	0.8503
Stay duration	The difference (in days) between the date of return (if any) and the departure date	114.08	160.81	114.30	160.93	0.8803
Days to departure	The difference (in days) between the date of departure and the date of the search	50.26	65.05	49.99	65.72	0.6520
Menu size	The number of itineraries found by the search engine	38.64	45.50	38.18	43.08	0.2484
Fastest duration	The duration (in hours) of the fastest itinerary proposed by the search engine	13.75	12.55	13.78	12.70	0.8101
Cheapest duration	The duration (in hours) of the cheapest itinerary proposed by the search engine	23.90	19.86	23.83	20.10	0.6848
Fastest price	The price (in USD) of the fastest itinerary proposed by the search engine	950.82	1426.07	934.12	1412.96	0.1927
Cheapest price	The price (in USD) of the cheapest itinerary proposed by the search engine	543.30	705.71	537.67	754.36	0.3895

Table 1: Mean and standard deviation of the session characteristics for both versions of the website.
Note: p-values indicate the p-value for a z-test comparing the proportion of (highlighted) conversion for each characteristic between the two versions of the website.

Results

At least one itinerary was selected in 10 877 of the 146 851 search sessions and at least one of the recommended itineraries was selected in 2 789 of the search sessions. We were able to identify requests made from the same web browser and, although we do not have the means to identify or track the person actually making the searches, we therefore consider searches on the same browser to be made by the same user. We identified 44 255 "users" in this way. Recurring users were always presented with the same version of the website (i.e., with either 2 or 3 highlighted itineraries), and conversions can thus be analyzed at the user level: 7 869 users converted at least once (17.78%), and 2 386 users chose a highlighted option at least once (5.39%).

Main effect

To assess statistical significance, we conducted two-sample, two-tailed z-tests and report Cohen's d as a measure of effect size. We also report Bayes Factor (BF_{01}) in favor of the null hypothesis.[8] We find that highlighting a decoy does not increase the probability that a user will convert, that is, choose at least one option in at least one of their search sessions (CR= 16.92% for the users presented with the NoDecoy version vs 17.26% for the Decoy version [z= 0.978, p= 0.328, Cohen′s d=0.00917, $BF_{01} = 46.73$]). Nor does it increase the probability that a user will select one highlighted option at least once (HCR(NoDecoy) = 5.04% vs HCR(Decoy) = 5.15% [z=0.501, p=0.617, Cohen's d=0.0047, $BF_{01} = 66.49$]). To investigate the targeted conversion rate, we focus our analysis on the search sessions within the Decoy version of the website in which decoys targeting both the Cheapest and the Fastest itineraries were available, and thus had equal chance of being highlighted. This condition is necessary for the targeted option to be randomly balanced between the Cheapest and the Fastest options and with respect to other attributes. We find that the probability of a user choosing the target option is not higher than the probability of choosing the highlighted competitor option (CR(target) = 2.08% vs CR(competitor) = 2.03%, z=−0.27, p= 0.788, Cohen′s d=0.00344, $BF_{01} = 39.23$).

[8] Bayes Factor t-tests are performed following Rouder et al. (2009). For each test, we report the "Scaled-Information Bayes Factor", based on a normal prior distribution of effect size, with scale factor r=0.70.

Test of moderators of the effect

To investigate whether the AE is more likely to appear in particular settings, we conduct a similar analysis, based on subsample of the sessions, splitting according to: (a) the menu size, (b) the time pressure (measured by the number of days before the departure), (c) the similarity between the Cheapest and the Fastest highlighted itineraries, (d) the economic stakes (measured by the price of the Cheapest itinerary), and (e) the device used (either mobile or computer). More precisely, for each of the following characteristics: size of the menu, time pressure, duration similarity, price similarity, and economic stakes, we separated the sessions according to their median values:

- High (low) "Size of the menu" sessions correspond to the half of the sessions with the highest (lowest) number of available itineraries.
- High (low) "Time pressure" sessions correspond to the half of the sessions where the date of departure is the closest (farthest) from the date of the search.
- High (low) "Duration similarity" sessions correspond to the half of the sessions where the ratio of the durations between the Cheapest and the Fastest $\left(\frac{D_C}{D_F}\right)$ is the highest (lowest).
- High (low) "Price similarity" sessions correspond to the half of the sessions where the ratio of the prices between the Fastest and the Cheapest $\left(\frac{P_F}{P_C}\right)$ is the highest (lowest).
- High (low) "economic stakes" sessions correspond to the half of the sessions where the price (in USD) of the Cheapest option is the highest (lowest).

For each of those subsamples, we then conducted the AE analysis following the same methodology as described previously. We report those results in Table 2 (for the conversion rate and the highlighted conversion rate) and in Table 3 (for the Target and Competitor Conversion rate).

			NoDecoy	Decoy	z	p-value	Cohen's d	BF_{01}
All sessions		CR	17.26%	16.92%	-0.978	0.328	-0.00917	46.73
		HR	5.15%	5.04%	-0.501	0.617	-0.0047	66.49
Size of the menu	high	CR	15.23%	15.20%	-0.059	0.953	-0.00068	61.25
		HCR	2.42%	2.50%	0.415	0.678	0.00479	56.20
	low	CR	13.15%	13.77%	1.533	0.125	0.01796	18.64
		HCR	5.62%	5.76%	0.535	0.592	0.00627	52.30
Time pressure	high	CR	14.76%	14.93%	0.377	0.706	0.00473	52.52
		HCR	4.35%	4.40%	0.222	0.824	0.00279	55.01
	low	CR	17.03%	17.39%	0.75	0.453	0.00954	41.96
		HCR	4.99%	5.14%	0.553	0.581	0.00703	47.70
Duration similarity	high	CR	15.08%	14.94%	-0.304	0.761	-0.00404	50.78
		HCR	4.76%	4.60%	-0.563	0.573	-0.00749	45.39
	low	CR	14.06%	14.48%	1.136	0.256	0.01215	34.68
		HCR	3.78%	3.94%	0.814	0.415	0.00871	47.47
Price similarity	high	CR	14.12%	14.03%	-0.196	0.844	-0.00253	53.83
		HCR	4.98%	5.01%	0.088	0.93	0.00114	54.66
	low	CR	14.43%	15.02%	1.544	0.123	0.01662	19.96
		HCR	3.42%	3.62%	0.993	0.321	0.01068	40.14
Economic Stakes	high	CR	15.03%	14.97%	-0.141	0.888	-0.0018	55.01
		HCR	4.09%	4.0%	-0.350	0.726	-0.00445	52.26
	low	CR	15.42%	15.98%	1.293	0.196	0.01547	25.62
		HCR	4.79%	5.03%	0.917	0.359	0.01097	38.82
Device	Computer	CR	17.60%	18.02%	0.959	0.338	0.01105	38.74
		HCR	5.34%	5.32%	-0.071	0.943	-0.00082	61.20
	Mobile	CR	13.85%	13.97%	0.180	0.857	0.00339	37.03
		HCR	3.61%	3.93%	0.868	0.385	0.01632	25.83

Table 2: The Attraction Effect on the overall and highlighted conversion rates separated by session characteristics.

Note: "CR" (respectively "HCR") indicates the overall (resp. highlighted) conversion rate, that is the probability of a user selecting at least one option (resp. to select the Fastest or the Cheapest highlighted itinerary) in at least one search session.

		Competitor	Target	z	p-value	Cohen's d	BF_{01}
All sessions		2.08%	2.03%	−0.27	0.788	0.00344	39.23
Size of the menu	high	1.14%	1.00%	-0.894	0.371	-0.01372	21.87
	low	2.48%	2.75%	0.95	0.342	0.01693	17.88
Time pressure	high	2.03%	1.63%	-1.78	0.075	-0.03034	6.03
	low	1.72%	2.26%	2.197	0.028	0.0385	2.56
Duration similarity	high	2.37%	2.04%	-1.111	0.266	-0.02268	13.23
	low	1.45%	1.64%	1.107	0.268	0.01592	18.87
Price similarity	high	2.10%	2.10%	0	1	0	27.91
	low	1.40%	1.51%	0.575	0.566	0.00879	27.71
Economic Stakes	high	2.18%	2.28%	0.444	0.657	0.00729	27.66
	low	1.41%	1.38%	-0.151	0.880	-0.00268	27.81
Device	Computer	2.14%	2.28%	0.626	0.531	0.00959	37.94
	Mobile	1.51%	1.41%	-0.320	0.749	-0.00821	37.06

Table 3: The Attraction Effect on target and competitor conversion rates.

Note: The target (competitor)'s conversion rate represents the probability of a user choosing the highlighted alternative (not) targeted by the decoy, in at least one of the search sessions.

We found no significant AE effect on conversion rate, highlighted conversion rate nor on target vs competitor conversion rate for any of the subpopulations, except a very small and limited effect in half of the sessions with the lowest time pressure (i.e. sessions where the departure date is more than 28 days ahead from the search). In those sessions, the alternatives are more likely to be chosen when they are targeted by a decoy (CR(Target)=2.26%), than when a decoy is targeting another alternative (CR(Comp)=1.72%). [Table 3, z=2.197, p=0.028, Cohen's d=0.0385, $BF_{01} = 2.56$].

Robustness test

As a robustness check, we conducted four logistic regressions at the session level, with the overall conversion rate (CR, Model 1), the highlighted conversion rate (HCR, Model 2), the conversion rate of the cheapest (CR(Cheapest), Model 3) and of the fastest (CR(Fastest), Model 4) as dependent variables. To limit the endogeneity issue, we only selected the first search session for each user. For Model 1 and Model 2, the AE effect was estimated using a dummy indicating if a decoy was highlighted. For Model 3 (Model 4), the AE was estimated using a dummy indicating the presence of the decoy targeting the Cheapest (Fastest) and the analysis was restricted to the subset of sessions where a decoy was highlighted and where decoys targeting both cheapest and fastest itineraries were available. Indeed, those conditions are necessary for the target to be exogenously and randomly decided with probability one half. The AE estimated in Model 3 and 4 therefore measures the impact of being the target vs being the competitor. In all the models, we use the abovementioned session characteristics (size of menu, Day to departure, Price similarity, Duration similarity and Used device), as moderators of the AE, as well as other session characteristics (price and duration of the fastest and cheapest itineraries, a dummy indicating if the itinerary is one way or return, and the duration of the stay for the return itineraries) as controls. The results are presented in Table 4.

	CR (Model 1)	HCR (Model 2)	CR(Cheapest) (Model 3)	CR(Fastest) (Model 4)
AE	-0.001	-0.012	0.259	-0.107
	(0.041)	(0.099)	(0.311)	(0.454)
AE x Days to departure	0.012	-0.034	0.015	-0.079
	(0.034)	(0.065)	(0.155)	(0.312)
AE x Menu size	0.003	0.209	-0.637	0.498
	(0.034)	(0.166)	(0.414)	(0.345)
AE x Duration similarity	-0.060	-0.157	0.862	-1.264
	(0.043)	(0.081)	(0.522)	(0.706)
AE x Price similarity	0.291	0.424	2.439	-3.062
	(0.304)	(0.936)	(2.615)	(3.909)
AE x Mobile Device	0.024	0.138	-0.054	-0.885
	(0.083)	(0.155)	(0.425)	(0.671)
Constant	-1.430***	-2.821***	-3.605***	-4.944***
	(0.104)	(0.208)	(0.667)	(0.942)
Days to departure	-0.091***	0.021	0.025	0.060
	(0.027)	(0.052)	(0.108)	(0.236)
Menu size	0.011	-0.881***	-0.687**	-0.401
	(0.026)	(0.137)	(0.235)	(0.332)
Duration similarity	-0.116**	-0.033	-1.490**	-0.320
	(0.037)	(0.062)	(0.511)	(0.372)
Price similarity	-1.508***	-4.025***	-4.015	0.123
	(0.298)	(0.886)	(2.678)	(2.192)
Mobile Device	-0.280***	-0.434***	-0.688*	0.553
	(0.066)	(0.123)	(0.292)	(0.438)
One-way itinerary	-2.473***	-2.031***	-1.478	-2.892
	(0.255)	(0.416)	(1.062)	(1.642)
Stay duration	1.133***	0.937***	0.615	1.370
	(0.111)	(0.180)	(0.459)	(0.712)
Price fastest	0.0001*	-0.0001	0.00002	-0.0003
	(0.00003)	(0.0001)	(0.0003)	(0.0004)
Price cheapest	-0.293***	-0.143	-0.567	0.385
	(0.053)	(0.118)	(0.361)	(0.392)
Duration fastest	0.038	0.214**	0.554*	-0.926
	(0.046)	(0.083)	(0.248)	(0.515)
Duration cheapest	-0.002	-0.010*	-0.009	0.028
	(0.002)	(0.004)	(0.014)	(0.024)
Observations	35,126	35,126	6,279	6,279
LL	-11,999.560	-4,596.985	-681.033	-256.951
AIC	24,035.120	9,229.971	1,398.067	549.901

Table 4: Regression analysis

*Note: standard errors are in parenthesis. *: p<0.05; **: p<0.01; ***: p<0.001*

Exploratory analysis

As exploratory analyses, we tested other cut-off values for the characteristics presented in Tables 2 and 3, and conducted similar analyses for other subpopulations. For each of the selected characteristics (size of the menu, time pressure, duration similarity, price similarity, and economic stakes), we divided the search sessions into subpopulations corresponding to deciles of those characteristics, and estimated the effect size of the AE on conversion rate, highlighted conversion rate and target vs competitor conversion rates. The detailed results are presented in Appendix B (Figures B1-B5). We also performed an analysis by country, restricted to the countries where we observed most of our search sessions (Appendix B, Figure B6). Since we conducted a large number of tests and only a very limited fraction of them reached significance, we will refrain from overinterpreting the significant results that are reported here.

Concerning time pressure, we found no significant effect when dividing the sessions into deciles (Figures B1 (a-c). However, we found opposite significant effects of highlighting a decoy on the probability of choosing the target compared to the competitor, when search sessions were divided according to being made a day, a week, or a month before travel. More precisely, the alternatives targeted by the decoy have less chance of being selected when the search is made between a week and a month before travel [Figure B1 (d), Cohen's d=-0.048, p=0.02], and more chance of being selected when the search session is made more than a month before the travel [Figure B1 (d) Cohen's d=0.036, p=0.045]. We found no impact of the menu size (Figure B2).

Concerning duration similarity, we found a negative effect on the overall conversion rate for the sessions corresponding to the first decile in terms of duration similarity $(D_C \geq 4.28 \times D_F)$ [Figure B3 (a) Cohen's d = -0.051, p=0.022]. For the sessions corresponding to the 7th decile $(1.33 \times D_F \leq D_C < 1.51 \times D_F)$ we found a positive effect on the probability of choosing a highlighted option [Figure B3 (b) Cohen's d = 0.052; p=0.014], however the probability of choosing the target is lower than the probability to choose the competitor [Figure B3 (c): Cohen's d= -0.058, p= 0.05].

As for price similarity, highlighting a decoy has a significant effect only if we restrict the analysis to the 7th decile $(1.25 \times P_C \leq P_F \leq 1.35 \times P_C)$, and reduces the overall conversion rate [Figure B4 (a): Cohen's d = -0.041; p=0.044]. Concerning economic stakes, when looking at Figure B5a, the effect size of the impact of highlighting a decoy on the probability of choosing any alternative seems to increase between the 4th to the 10th decile, although none of the effects is statistically significant.

Finally, concerning countries' subpopulations, the only significant effect was obtained for the sessions made from Germany: highlighting a decoy increases the probability of choosing the target, compared to the competitor [Figure B6 (c): Cohen's d = 0.129, p=0.004].

Discussion

In this article, we studied and did not find evidence for the existence of the attraction effect (AE) for the recommended options in online flight booking. In general, our data did not provide support for the following three hypotheses: *H1: recommending a decoy option increases the probability that the user will choose one of the available options out of all options available on the platform; H2: recommending a decoy option increases the probability that the user will choose one of the recommended options; H3: recommending a decoy option increases the probability that the user will choose the targeted option (i.e., a recommended option that dominates the decoy).*

The typical paradigm used in the studies documenting AE is based on hypothetical forced choices among two or three alternatives. These alternatives are usually characterized by two or three numerical attributes and presented in a simple, minimalistic design. In contrast, our flight aggregator offers the users the possibility to search and gather information about numerous alternatives, defined by numerical attributes for which preferences are universal (such as price and flight duration), as well as ones for which preferences could be potentially heterogenous across users (e.g., departure time or airline). In addition, implementation in the field cannot prevent users from soliciting additional information from other search engines. In this setting, we decided to manipulate which options were presented in the highlighted panel at the top of the list of all available alternatives. Because of its saliency, we considered that this highlighted panel constitutes (to some extent) a refined design, embedded in a more complex environment. We, therefore, expected to detect an AE, but at somewhat weaker level. Potential difficulties in identifying dominance relationships could be attributed to the presence of certain flight attributes that obscure the user search process. Moreover, the richness of the choice set, along with the option to completely abstain from choosing, repeat the search at a later stage, or use another flight aggregator, could also influence the construction of user preferences. Our manipulation, therefore, represents a relatively less pronounced modification of the user experience, compared with previous (laboratory) studies. In particular, a user who repeated their search multiple times over an extended period and investigated extensive set of alternatives, would form much more precise and stable preferences than in a more artificial laboratory setting, significantly diminishing (and potentially annihilating) the effect of the nudge. One would expect this limiting effect on the AE to be smaller in users' early searches. However, our robustness analyses did not detect the AE in the subset of data consisting of the search sessions by novice users either (Table 4). In addition to the points discussed in the previous paragraph, there are a number of other possible explanations for the lack of AE observed in our setting in contrast to previous studies.

First, in our experiment, decoy alternatives were included in the choice set in both versions of the website and our manipulation consisted only in recommending the decoys in one of the versions. Therefore, if users were attentive and considered all the available alternatives, the choice set was similar in both versions of the website, preventing the AE from occurring. However, this interpretation seems unlikely since search sessions returned on average more than 38 possible itineraries (preventing users from full consideration). Moreover, when controlling for the size of the choice set, we find no evidence for the effect, regardless of the number of available alternatives. This suggests that the number of available alternatives was not the main reason why we were unable to replicate the effect.

Another explanation could be that most consumers have strong preferences toward one of the proposed itineraries, or toward the outside option. The AE has mostly been observed when the target and the competitor were close in terms of utility. However, our analyses do not suggest that a lack of similarity in terms of utility was the main reason for not finding the effect. Indeed, we were not able to observe any monotonic relation between the similarity in terms of price and duration of the competitor and the target and the strength of the effect. If similarity does play a role in the field, its role is thus less straightforward than the one identified in laboratory settings.

Another particularity of flight booking decisions is to involve stakes much higher than in most incentivized experiments, leading to more deliberative thinking and preventing the appearance of the AE (Gaudeul & Crosetto, 2019). However, we conducted analyses on subpopulations to control for economic stakes, measured by the price of the cheapest of the available alternatives, and did not find that the stakes of the decisions were moderating the effect in the field.

Another reason for the field replication failure could be that users are less captive than in laboratory experiments, in the sense that they have access to richer outside options (other platforms, possibility to delay their choices, etc.) Unfortunately, we were not able to measure user's activity outside of the flight aggregator and cannot directly test the captivity hypothesis (e.g., by controlling the number of websites visited). The closest proxy of captivity we could get from our data was time pressure, measured by the delay between search and travel. Indeed, one may assume the richness of the outside options decreases as the departure draws nearer, increasing both user captivity and time pressure. Interestingly, we observed a positive (although very small in size) attraction effect, only when we restricted our analyses to the half of the sessions where search was the most in advance of the travel. While this result goes against previous findings, which suggest that time pressure enhances AE (Gaudeul & Crosetto, 2019), and thus should be interpreted carefully, at least it does not suggest that captivity is a necessary condition for obtaining the effect.

Finally, one crucial difference is that in our experiment, the recommended decoy was chosen from among the set of potential decoys. Other studies instead created and offered a specific decoy, that was expected (by the choice architect) to maximize the probability of observing the AE. Indeed, Kaptein et al. (2016) have shown that the size of the AE is a (non-monotonic) function of the distance between the decoy, the target and the competitor, and have proposed an algorithm to find the decoy that maximizes the AE. They conclude that the failure to replicate the AE can be attributed to a "misplaced" decoy. Applying their algorithm in our case was not possible, due to the heterogeneity of the sessions and to the fact that we cannot freely manipulate the price and duration of itineraries. This absence of latitude is shared by many recommender systems: the choice architect does not always have the ability (or the right) to create new alternatives or to alter existing ones. This last explanation suggests that digital decoy nudges could be more suited to situations where the choice architect has the flexibility to create and modify the proposed alternatives (e.g., a producer or a reseller, who can decide the price or other attributes of products), rather than those where the main strategic variable is the decision of which and how existing information is disclosed (e.g., a search engine or a marketplace).

On the other hand, additional exploratory analyses have suggested possible heterogeneity in the AE.

This conjecture, however, needs to be confirmed by confirmatory research based on new samples (due to the exploratory nature of our analyses). In particular, we found a significant effect only for very specific levels of similarity, countries, and time pressure. Because a high number of statistical tests were performed to find those results, it is likely they were obtained by chance, so we refrain from interpreting them. Nevertheless, being able to find (some) effect heterogeneity in subpopulations suggests that it could be possible to design algorithms that learn to identify which session features support the effect the most and thus offer personalized digital nudges. For example, one of our exploratory results suggested that opposite effects could be found depending on how far in advance of the travel date the flight search was conducted. Recommending a decoy favored the target alternatives (as predicted by AE) when the search preceded the travel by more than a month, but the decoy favored the competing alternative (repulsion effect) when the search preceded the travel by a shorter period. If this finding were robust, a recommender system could be designed to exploit this feature and adapt digital nudges to session characteristics. For example, if the goal of the flight aggregator presented here were to promote faster itineraries over cheaper ones, it could recommend decoys that target the fastest when the day of departure is more than a month in the future, but recommend decoy options that target the cheapest for searches preceding the departure by shorter periods.

To conclude, more research is needed to confirm these findings. In particular, we call for more conceptual replications of the AE within real digital marketplaces and recommender systems and with a larger variety of goods. If other researchers fail to replicate the AE in recommender systems, then laboratory experiments could also be useful to identify and test how the specific features of recommender systems can moderate the AE. For instance, one could test how the recommendation model (e.g., recommending options based on compensatory versus non-compensatory models (Lin et al., 2019)) or the interface (allowing or not the filtering of the alternatives with respect to attributes) of the recommender system interacts with the AE. Another interesting feature to explore would be the absence of captivity of recommender system users who are free to search on other websites. For instance, one could imagine an experiment where the participants have access to several platforms from which they can make their choice, and the study investigates whether the presence (or recommendation) of a decoy option on one platform, but not on the others, impacts how choices are made between and within platforms. Such studies would not only guide managers in devising more efficient recommender systems, but also contribute to advancing our understanding of human decision-making in digital environments.

References.

Adomavicius, G., Bockstedt, J. C., Curley, S. P., & Zhang, J. (2013). Do recommender systems manipulate consumer preferences? A study of anchoring effects. *Information Systems Research*. https://doi.org/10.1287/isre.2013.0497

Adomavicius, G., Bockstedt, J. C., Curley, S. P., & Zhang, J. (2018). Effects of online recommendations on consumers' willingness to pay. *Information Systems Research*. https://doi.org/10.1287/isre.2017.0703

Adomavicius, G., Bockstedt, J. C., Curley, S. P., & Zhang, J. (2019). Reducing recommender system biases: An investigation of rating display designs. *MIS Quarterly: Management Information Systems*. https://doi.org/10.25300/MISQ/2019/13949

Aerts, G., Smits, T., & Verlegh, P. W. J. (2017). The platform shapes the message: How website design affects abstraction and valence of online consumer reviews. *Decision Support Systems*. https://doi.org/10.1016/j.dss.2017.10.006

Arnott, D., & Gao, S. (2019). Behavioral economics for decision support systems researchers. *Decision Support Systems*. https://doi.org/10.1016/j.dss.2019.05.003

Arnott, David, & Gao, S. (2019). Behavioral economics for decision support systems researchers. *Decision Support Systems*, *122*(February), 113063. https://doi.org/10.1016/j.dss.2019.05.003

Babutsidze, Z. (2012). How do consumers make choices? a survey of evidence. *Journal of Economic Surveys*. https://doi.org/10.1111/j.1467-6419.2011.00693.x

Banerjee, P., Chatterjee, P., Mishra, S., & Mishra, A. A. (2020). When should we not expect attraction effect? The moderating influence of analytic versus holistic thinking. *Journal of Strategic Marketing*, *28*(5), 399–416. https://doi.org/10.1080/0965254X.2018.1511629

Bettman, J. R., Luce, M. F., & Payne, J. W. (1998). Constructive consumer choice processes. *Journal of Consumer Research*, *25*(3), 187–217.

Bobadilla, J., Ortega, F., Hernando, A., & Gutiérrez, A. (2013). Recommender systems survey. *Knowledge-*

Based Systems, *46*, 109–132. https://doi.org/10.1016/j.knosys.2013.03.012

Brady, R. L., & Rehbeck, J. (2016). Menu-Dependent Stochastic Feasibility. *Econometrica*, *84*(3), 1203–1223.

Camilleri, A. R. (2020). The importance of online reviews depends on when they are presented. *Decision Support Systems*, *133*(June 2019), 113307. https://doi.org/10.1016/j.dss.2020.113307

Castillo, G. (2020). The attraction effect and its explanations. *Games and Economic Behavior*, *119*, 123–147.

Choplin, J. M., & Hummel, J. E. (2002). Magnitude comparisons distort mental representations of magnitude. *Journal of Experimental Psychology: General*, *131*(2), 270–286. https://doi.org/10.1037/0096-3445.131.2.270

Craciun, G., Zhou, W., & Shan, Z. (2020). Discrete emotions effects on electronic word-of-mouth helpfulness: The moderating role of reviewer gender and contextual emotional tone. *Decision Support Systems*. https://doi.org/10.1016/j.dss.2019.113226

Dasgupta, I., & Pattanaik, P. K. (2007). 'Regular'choice and the weak axiom of stochastic revealed preference. *Economic Theory*, *31*(1), 35–50.

Dhar, R., & Simonson, I. (2003). The effect of forced choice on choice. *Journal of Marketing Research*, *40*(2), 146–160.

Dimara, E., Bezerianos, A., & Dragicevic, P. (2017). The attraction effect in information visualization. *IEEE Transactions on Visualization and Computer Graphics*, *23*(1).

Doyle, J. R., O'Connor, D. J., Reynolds, G. M., & Bottomley, P. A. (1999a). The robustness of the asymmetrically dominated effect: Buying frames, phantom alternatives, and in-store purchases. *Psychology and Marketing*, *16*(3), 225–243. https://doi.org/10.1002/(SICI)1520-6793(199905)16:3<225::AID-MAR3>3.0.CO;2-X

Doyle, J. R., O'Connor, D. J., Reynolds, G. M., & Bottomley, P. A. (1999b). The robustness of the asymmetrically dominated effect: Buying frames, phantom alternatives, and in-store purchases. *Psychology and Marketing*, *16*(3), 225–243. https://doi.org/10.1002/(SICI)1520-6793(199905)16:3<225::AID-MAR3>3.0.CO;2-X

Frederick, S., Lee, L., & Baskin, E. (2014). The limits of attraction. *Journal of Marketing Research*, *51*(4), 487–507.

Gaudeul, A., & Crosetto, P. (2019). *Fast then slow: A choice process explanation for the attraction effect*.

Gomez, Y., Martínez-Molés, V., Urbano, A., & Vila, J. (2016). The attraction effect in mid-involvement categories: An experimental economics approach. *Journal of Business Research*, *69*(11), 5082–5088. https://doi.org/10.1016/j.jbusres.2016.04.084

Gupta, S., Kohavi, R., Tang, D., Xu, Y., Andersen, R., Bakshy, E., Cardin, N., Chandran, S., Chen, N., Dominic, C., Curtis, M., Deng, A., Duan, W., Forbes, P., Brian, F., Guy, T., Imbens, G. W., Guillaume, S., Jacques, … Igor, Y. (2019). Top Challenges from the first Practical Online Controlled Experiments Summit. *ACM SIGKDD Explorations Newsletter*, *21*(1), 20–35.

Ho, S. Y., & Lim, K. H. (2018). Nudging moods to induce unplanned purchases in imperfect mobile personalization contexts. *MIS Quarterly: Management Information Systems*. https://doi.org/10.25300/MISQ/2018/14083

Huber, J., Payne, J. W., & Puto, C. (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research*, *9*(1), 90–98.

Huber, J., Payne, J. W., & Puto, C. P. (2014). Let's be honest about the attraction effect. *Journal of Marketing Research*, *51*(4), 520–525.

Hüffmeier, J., Mazei, J., & Schultze, T. (2016). Reconceptualizing replication as a sequence of different studies: A replication typology. *Journal of Experimental Social Psychology*, *66*, 81–92. https://doi.org/10.1016/j.jesp.2015.09.009

Jameson, A., Berendt, B., Gabrielli, S., Cena, F., Gena, C., Vernero, F., & Reinecke, K. (2014). Choice architecture for human-computer interaction. *Foundations and Trends in Human-Computer Interaction*. https://doi.org/10.1561/1100000028

Kaptein, M. C., Van Emden, R., & Iannuzzi, D. (2016). Tracking the decoy: maximizing the decoy effect through sequential experimentation. *Palgrave Communications*, *2*, 16082.

Kim, S., & Hasher, L. (2005). The attraction effect in decision making: Superior performance by older adults. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *58*(1), 120–133. https://doi.org/10.1080/02724980443000160

Kroll, T., & Stieglitz, S. (2019). Digital nudging and privacy: improving decisions about self-disclosure in social networks *. *Behaviour and Information Technology*. https://doi.org/10.1080/0144929X.2019.1584644

Lee, L., & Anderson, R. (2009). *A Comparison of Compensatory and Non- Compensatory Decision Making Strategies in IT Project Portfolio Management*.

Li, M., Sun, Y., & Chen, H. (2019). Retraction of “The Decoy Effect as a Nudge: Boosting Hand Hygiene With a Worse Option.” *Psychological Science*, *30*(7), 1107–1107. https://doi.org/10.1177/0956797619858006

Lichters, M., Bengart, P., Sarstedt, M., & Vogt, B. (2017). What really matters in attraction effect research: when choices have economic consequences. *Marketing Letters*, *28*(1), 127–138.

Lichters, M., Sarstedt, M., & Vogt, B. (2015). On the practical relevance of the attraction effect: A cautionary note and guidelines for context effect experiments. *AMS Review*, *5*(1–2), 1–19.

Lin, C., Shen, X., Chen, S., Zhu, M., & Xiao, Y. (2019). Non-compensatory psychological models for recommender systems. *33rd AAAI Conference on Artificial Intelligence, AAAI 2019, 31st Innovative Applications of Artificial Intelligence Conference, IAAI 2019 and the 9th AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019*, 4304–4311. https://doi.org/10.1609/aaai.v33i01.33014304

Liu, Q. Ben, & Karahanna, E. (2017). The dark side of reviews: The swaying effects of online product reviews on attribute preference construction. In *MIS Quarterly: Management Information Systems*. https://doi.org/10.25300/MISQ/2017/41.2.05

Lu, J., Wu, D., Mao, M., Wang, W., & Zhang, G. (2015). Recommender system application developments: A survey. *Decision Support Systems*. https://doi.org/10.1016/j.dss.2015.03.008

Mao, W., & Oppewal, H. (2012). The attraction effect is more pronounced for consumers who rely on intuitive reasoning. *Marketing Letters*, *23*(1), 339–351. https://doi.org/10.1007/s11002-011-9157-y

Mishra, S., Umesh, U. N., & Stem, D. E. (1993). Antecedents of the Attraction Effect: An Information-Processing Approach. *Journal of Marketing Research*, *30*(3), 331. https://doi.org/10.2307/3172885

Pettibone, J. C., & Wedell, D. H. (2007). Testing alternative explanations of phantom decoy effects. *Journal of Behavioral Decision Making*, *20*(3), 323–341.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. In *Psychonomic Bulletin and Review*. https://doi.org/10.3758/PBR.16.2.225

Schneider, C., & Weinmann, M. (n.d.). *Digital Nudging – Influencing Choices by Using Interface Design*.

Schneider, C., Weinmann, M., Mohr, P. N. C., Brocke, J., Schneider, C., Weinmann, M., & Mohr, P. N. C. (2021). *When the Stars Shine Too Bright : The Influence of Multidimensional Ratings on Online Consumer Ratings When the Stars Shine Too Bright : The In fl uence of Multidimensional Ratings on Online Consumer Ratings*. *May*, 0–28.

Shafir, S., Waite, T. A., & Smith, B. H. (2002). *Context-dependent violations of rational choice in honeybees (Apis mellifera) and gray jays (Perisoreus canadensis)*. 180–187. https://doi.org/10.1007/s00265-001-0420-8

Shen, Y., Sun, H., Suang, C., & Chuan, H. (2020). Facilitating Complex Product Choices on E-commerce Sites : An Unconscious Thought and Circadian Preference Perspective. *Decision Support Systems*, *137*(August), 113365. https://doi.org/10.1016/j.dss.2020.113365

Sheng, H., & Joginapelly, T. (2012). Effects of Web Atmospheric Cues on Users' Emotional Responses in E-Commerce. *AIS Transactions on Human-Computer Interaction*. https://doi.org/10.17705/1thci.00036

Simonson, I., & Tversky, A. (1992). Choice in Context: Tradeoff Contrast and Extremeness Aversion. *Journal of Marketing Research*, *29*(3), 281. https://doi.org/10.2307/3172740

Sivan, L., Smith, M. D., & Telang, R. (2019). Do search engines influence media piracy? Evidence from a randomized field study. *MIS Quarterly: Management Information Systems*. https://doi.org/10.25300/MISQ/2019/14404

Slaughter, J. E., Sinar, E. F., & Highhouse, S. (1999). Decoy effects and attribute-level inferences. *Journal of Applied Psychology*, *84*(5), 823–828. https://doi.org/10.1037/0021-9010.84.5.823

Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth and happiness*.

Trueblood, J. S., Brown, S. D., Heathcote, A., Busemeyer, J. R., & Busemeyer, J. R. (2013). Not Just for Consumers: Context Effects Are Fundamental to Decision Making. *Psychological Science*, *24*(April), 901–908. https://doi.org/10.1177/0956797612464241

Tversky, A., & Kahneman, D. (1975). Judgment under uncertainty: Heuristics and biases. In *Utility, probability, and human decision making* (pp. 141–162). Springer.

Weinmann, M., Mishra, A., Kaiser, L., & vom Brocke, J. (2020). The Attraction Effect in Reward-Based Crowdfunding: Evidence from Four Experiments. *SSRN Electronic Journal*.

https://doi.org/10.2139/ssrn.3630514

Weinmann, M., Schneider, C., & Brocke, J. vom. (2016). Digital Nudging. *Business and Information Systems Engineering*. https://doi.org/10.1007/s12599-016-0453-1

Wessel, M., Adam, M., & Benlian, A. (2018). The impact of sold-out early birds on option selection in reward-based crowdfunding. *Decision Support Systems*, #pagerange#. https://doi.org/10.1016/j.dss.2018.12.002

Yang, S. (2013). *THE ATTRACTION EFFECT: AN OVERVIEW, ITS FRAGILITY, AND A META-ANALYSIS* (Issue May).

Yang, S., & Lynn, M. (2014). More evidence challenging the robustness and usefulness of the attraction effect. *Journal of Marketing Research*, *51*(4), 508–513.

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics*, *13*(1), 75–98. https://doi.org/10.1007/s10683-009-9230-z

Appendices

Appendix A: Impact of highlighting a third itinerary.

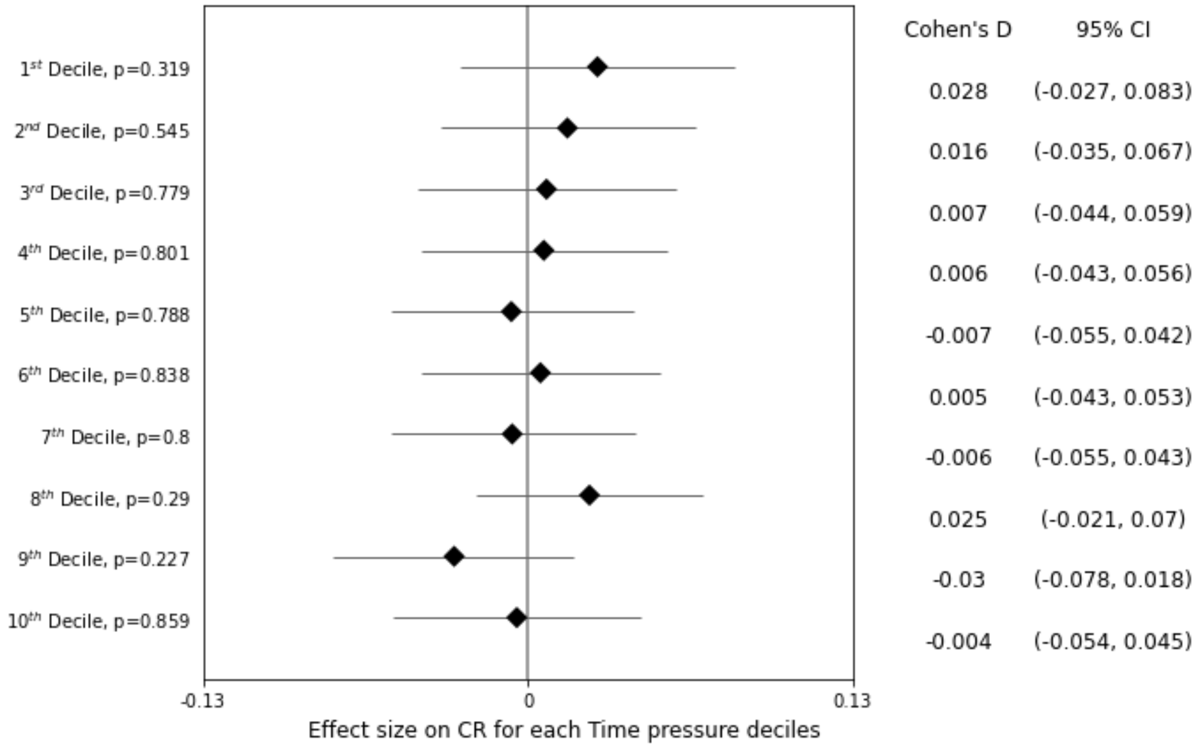
In this section, we address the question of the potential impact highlighting a third alternative could have on conversion rates.

In the main analysis of the AE on conversion rates, we excluded the 54,141 sessions where no potential decoys were available. Obviously, the sessions with no potential decoys (NPD sessions) are different from the sessions with potential decoys (PD sessions), in terms of characteristics (e.g., the more choices in a session, the higher likelihood to find potential decoys) and in terms of conversion (6.57% of the NPD sessions have a conversion vs 7.50% of the PD sessions [$z=3.92$, $p<0.0001$]).

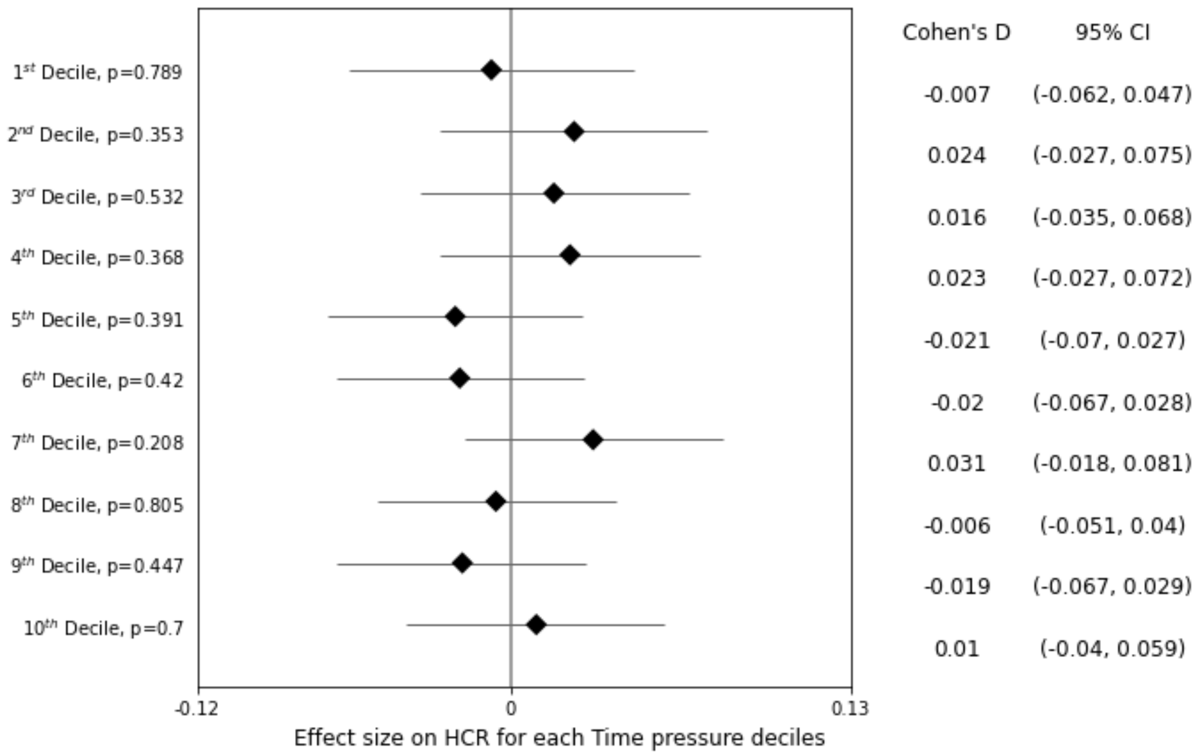
We therefore conduct an analysis of the conversion rate restricted to the NPD sessions, where the impact of randomly adding a third itinerary can be assessed. We found that adding a third highlighted option that is not a decoy does not statistically increase the overall conversion rate (CR(3H) = 13.41% vs CR(2H) = 12.91% [$z = 0.224$, $p = 0.823$, $d = 0.00225$]), nor the probability of choosing either the fastest or the cheapest recommended alternatives (HCR(3H) = 6.42 % vs HCR(2H) = 6.63 % [$z = -0.627$, $p=0.531$, Cohen's $d = -0.00857$]).

Appendix B: Exploratory analysis – Effect size analyzed for subpopulations.

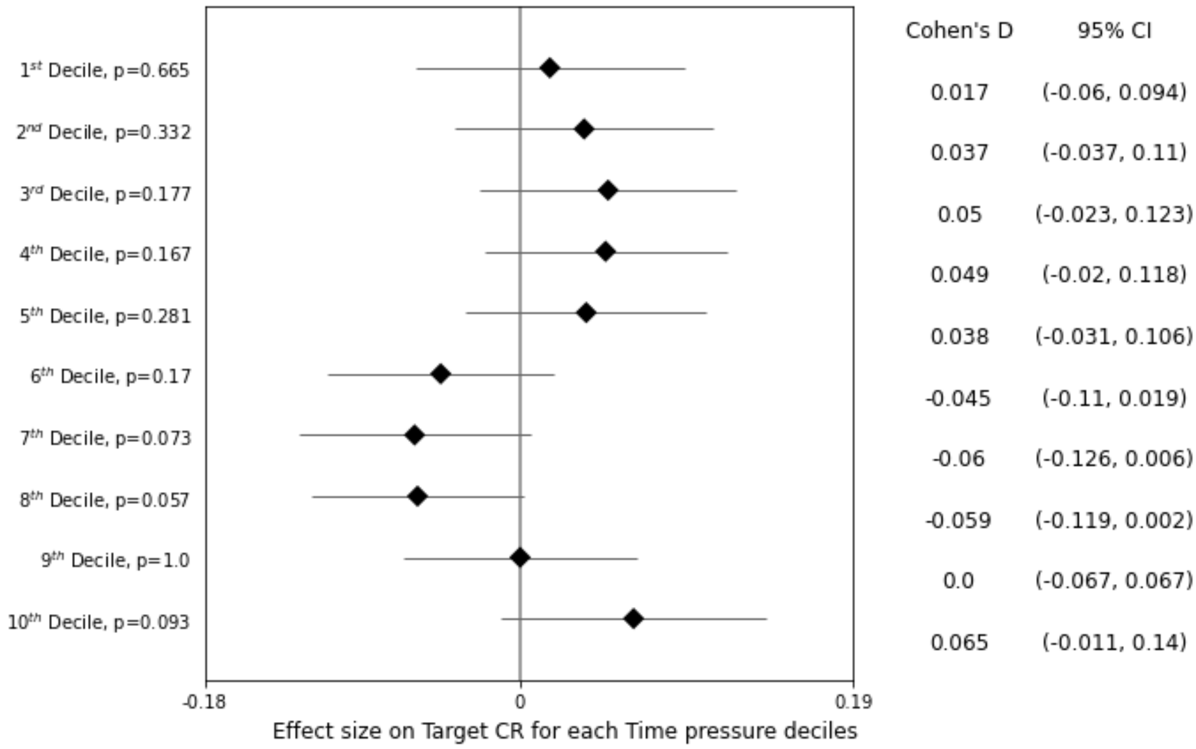
B-1) Time pressure



(a)



(b)



(c)

Figure B1 : Effect size estimated for each decile of Time pressure.
Note: Figures show respectively the effect size for (a) the overall conversion rate (CR), (b) the highlighted conversion rate (HCR), and (c) the difference between the target and the competitor conversion rate. Each line represents the effect size (Cohen's d), and 95% confidence interval estimated based on subsamples of the sessions that corresponds to each decile in terms of time pressure. The time pressure of a search sessions is defined by the number of days before the departure. The lower the number of days before departure, the higher the similarity (1st decile : days before departure ≥ 148, 2nd decile = [89, 148[, 3rd decile = [61, 89[, 4th decile = [41, 61[, 5th decile = [28, 41[, 6th decile = [18, 28[, 7th decile = [11, 18[, 8th decile = [6, 11[, 9th decile = [2, 6[, 10th decile days before departure < 2).

Figure B1 (c), suggests a common behavior between deciles 2-5 and between deciles 6-8. Therefore, we tested the attraction effect on the target and competitor conversion rates by spliting the sample according to values that correspond to those deciles. We decided to split the sample according to whether the departure date is (i) the same day as the search, (ii) less than a week ahead of the search, (iii) between a week and a month (30 days) ahead of the search, (iv) more than a month ahead of the search. Results for those particular subpopulations are presented in Figure B1 (d).

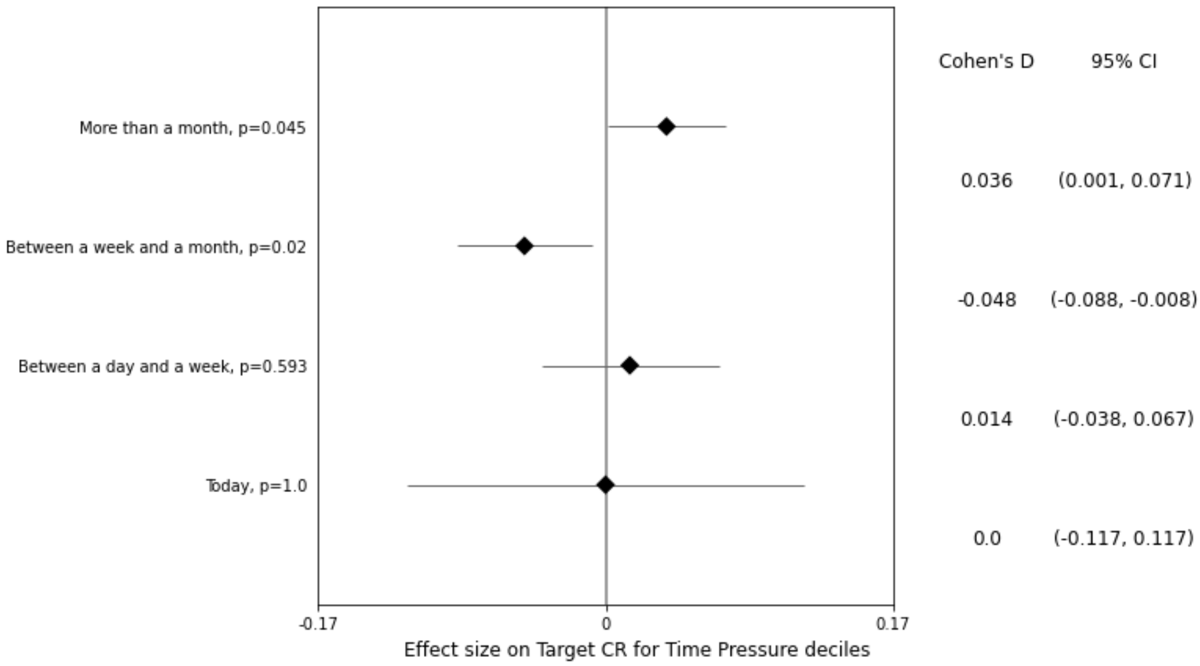
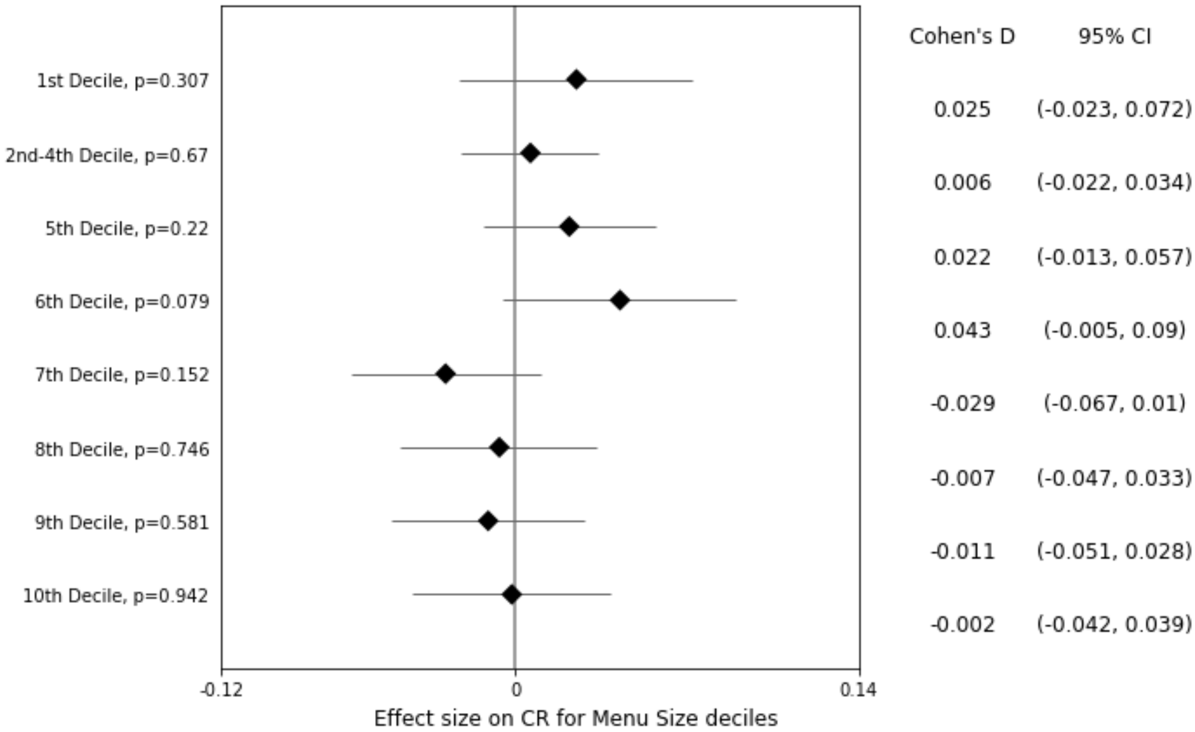
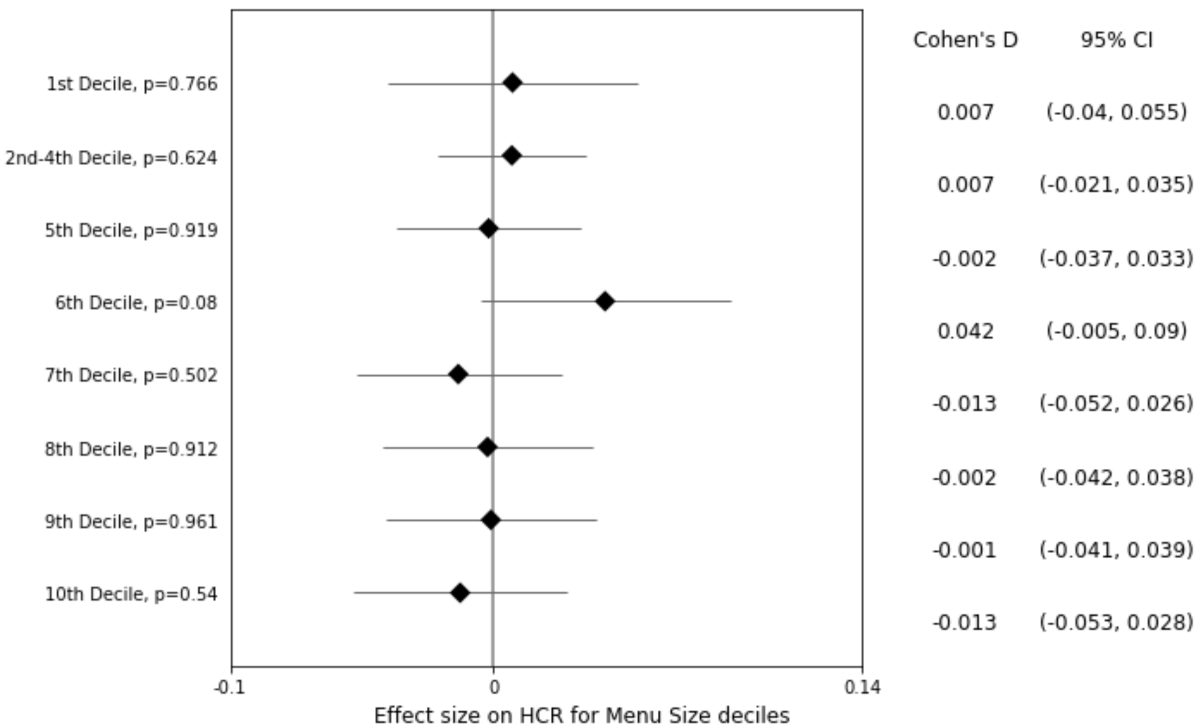


Figure B1 (d) : Effect size of Target vs Competitor CR as function of Time pressure.

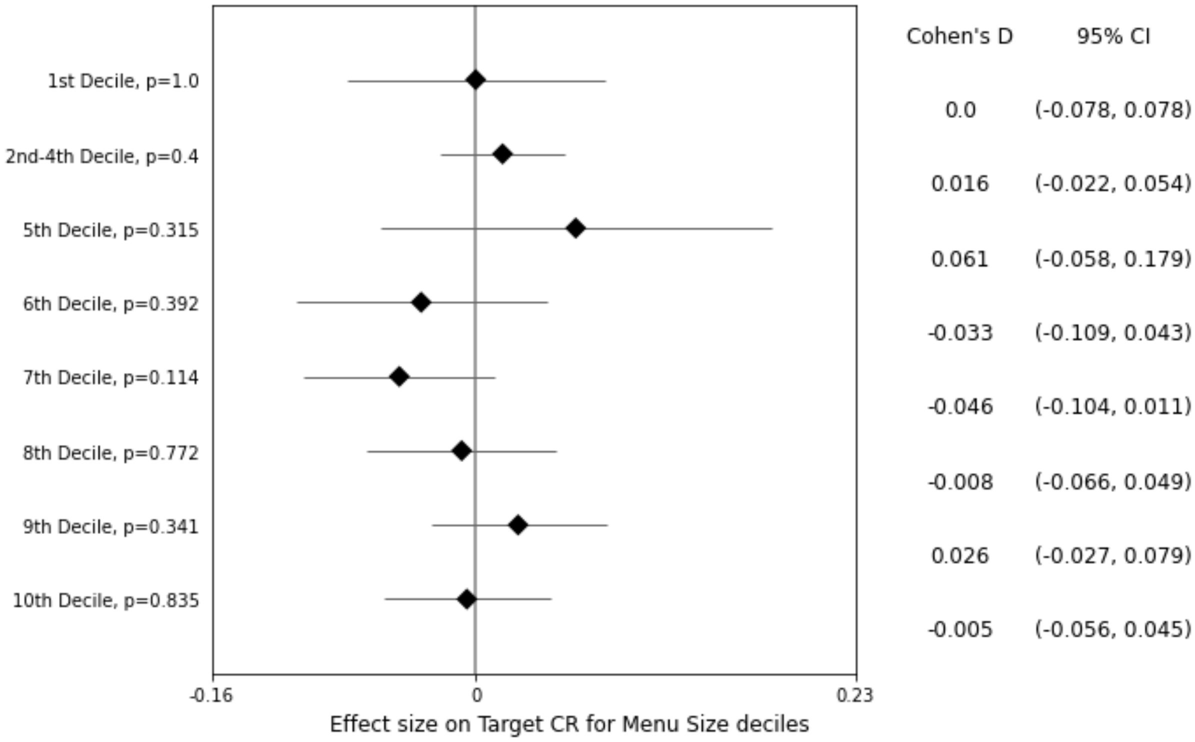
B-2) Menu size.



(a)



(b)

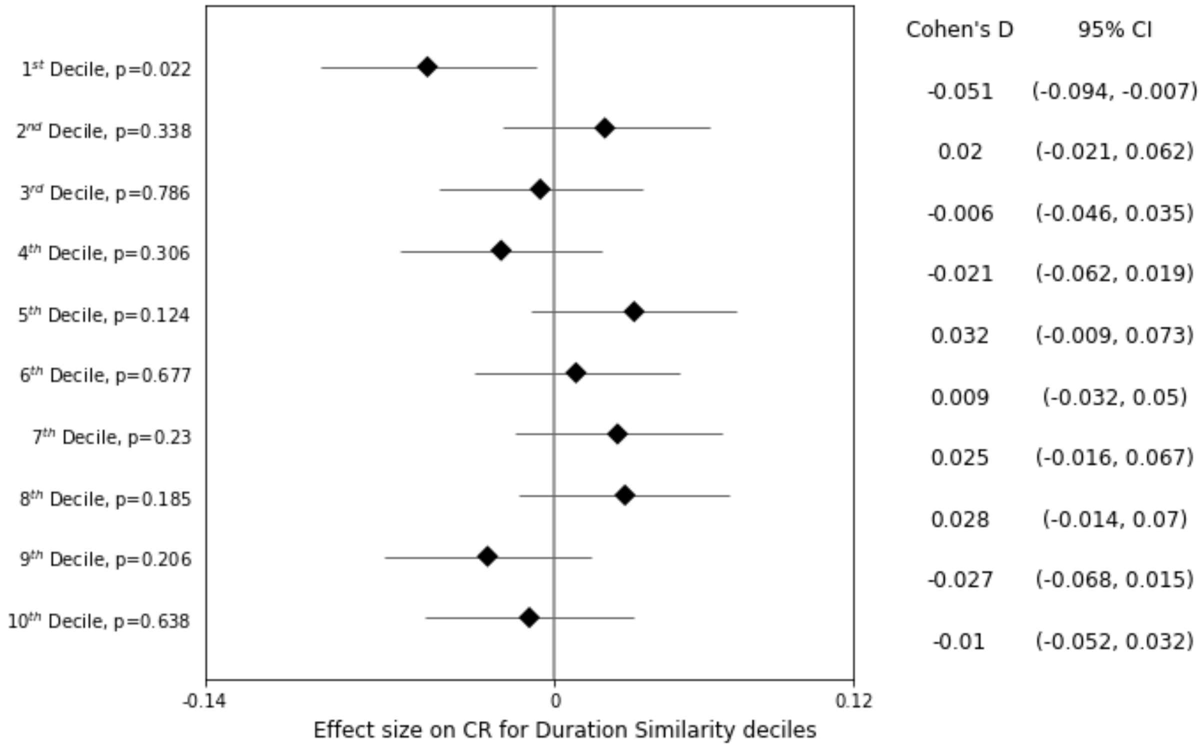


(c)

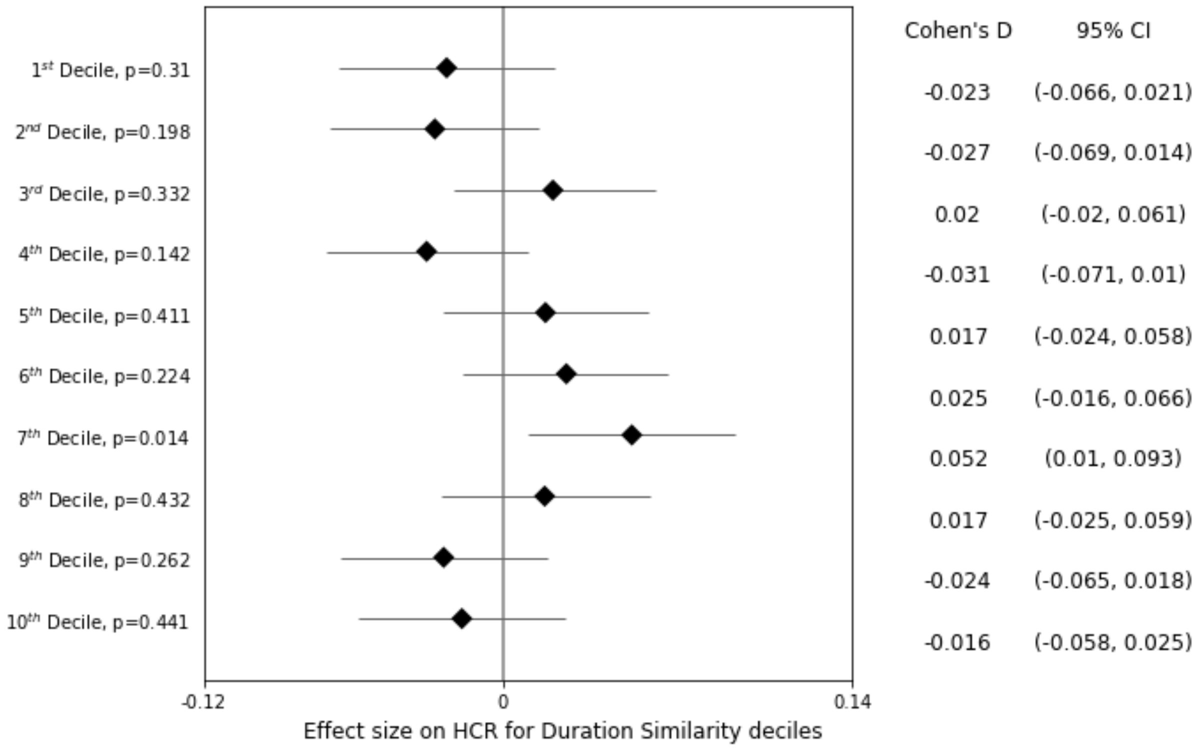
Figure B2 : Effect size estimated for each decile of Menu size

Note: Figures show respectively the effect size for (a) the overall conversion rate (CR), (b) the highlighted conversion rate (HCR), and (c) the difference between the target and the competitor conversion rate. Each line represents the effect size (Cohen's d), and 95% confidence interval estimated based on subsamples of the sessions that correspond to each decile in terms of Menu size. The menu size of a search session is defined by the number of alternatives disclosed to the user (1st decile : 3 ≤ number of alternatives < 20 2nd , 3rd and 4th deciles : 20 alternatives, 5th decile : 21 alternatives, 6th decile = [22, 30[, 7th decile = [30, 39[, 8th decile = [39, 50[, 9th decile = [50, 80[, 10th decile :≥ 80).

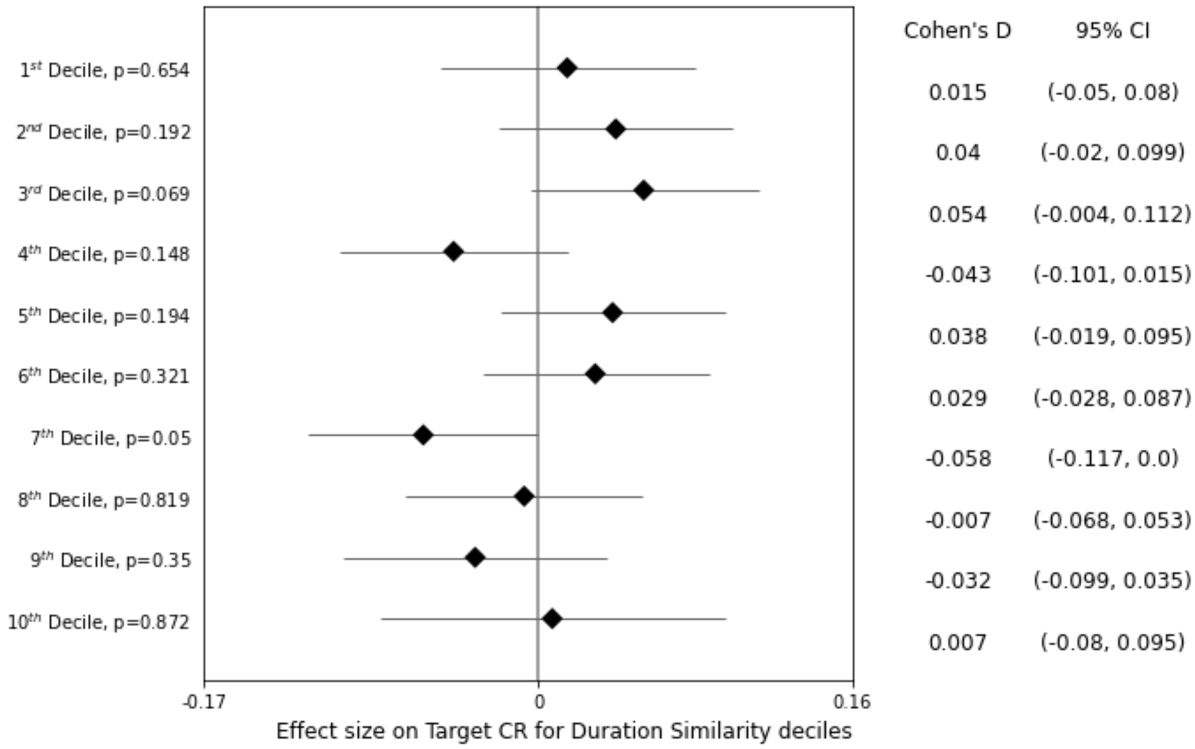
B-3) Duration similarity



(a)



(b)

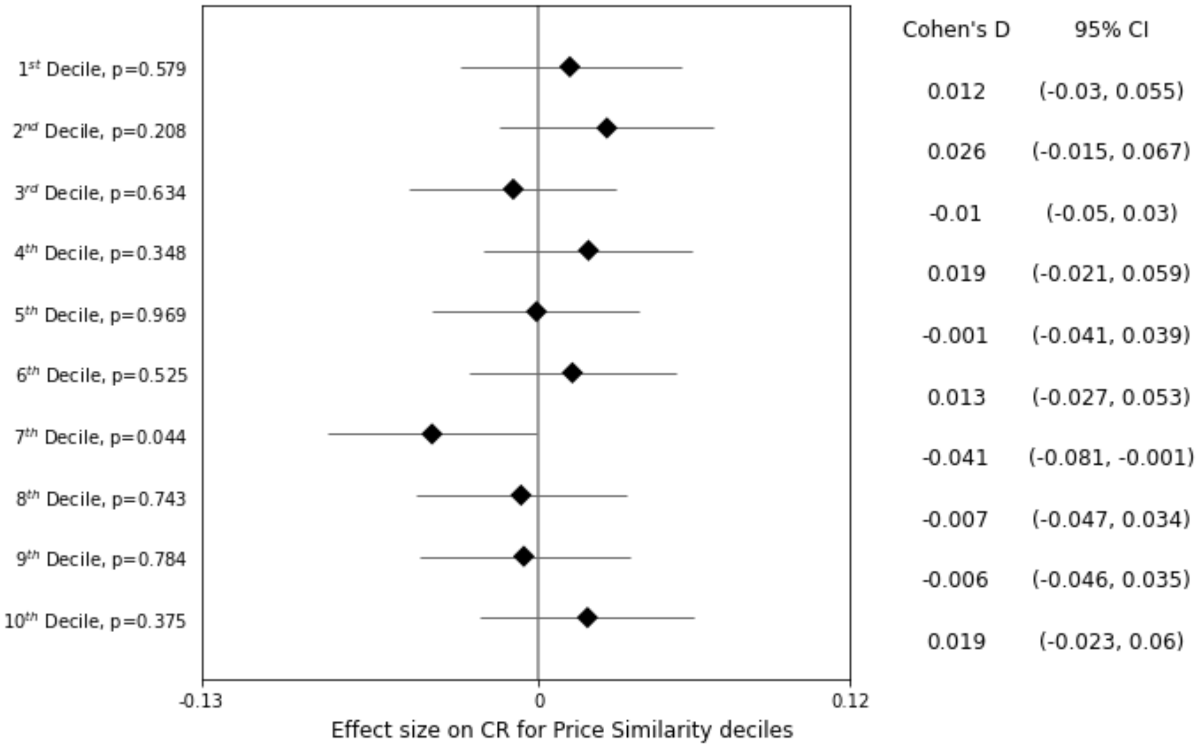


(c)

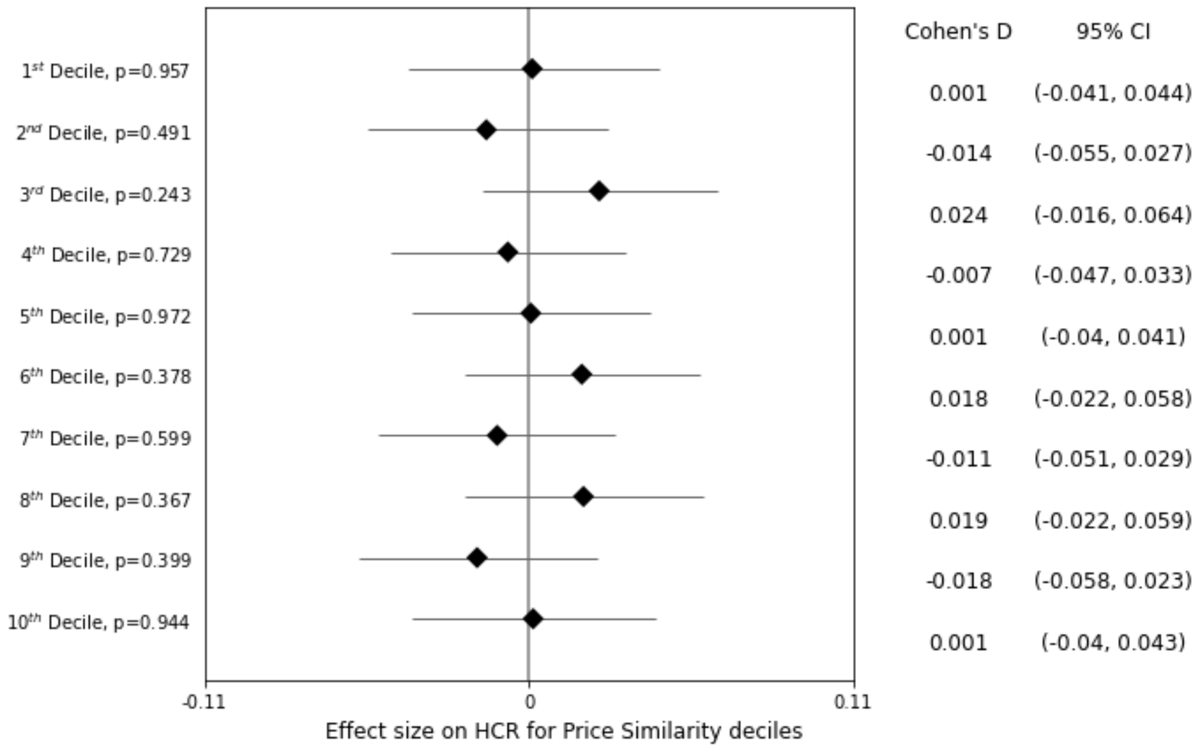
Figure B3 : Effect size estimated for each decile of duration similarity.

Note: Figures show respectively the effect size for (a) the overall conversion rate (CR), (b) the highlighted conversion rate (HCR), and (c) the difference between the target and the competitor conversion rate. Each line represents the effect size (Cohen's d), and 95% confidence interval estimated based on subsamples of the sessions that correspond to each decile in terms of duration similarity. The duration similarity of a search session is defined as the ratio between the duration of the Cheapest and the duration of the Fastest itineraries $\left(\frac{D_C}{D_F}\right)$. *The lower the ratio, the higher the similarity (1st decile:* $\frac{D_C}{D_F} \geq 4.28$ *; 2nd decile = [2.95,4.28[; 3rd decile = [2.33, 2.95[; 4th decile = [1.97, 2.33[; 5th decile = [1.71, 1.97[; 6th decile = [1.51, 1.71[; 7th decile = [1.33, 1.51[; 8th decile = [1.17, 1.33[; 9th decile = [1.07, 1.17[; 10th decile = [1, 1.07[ ).*

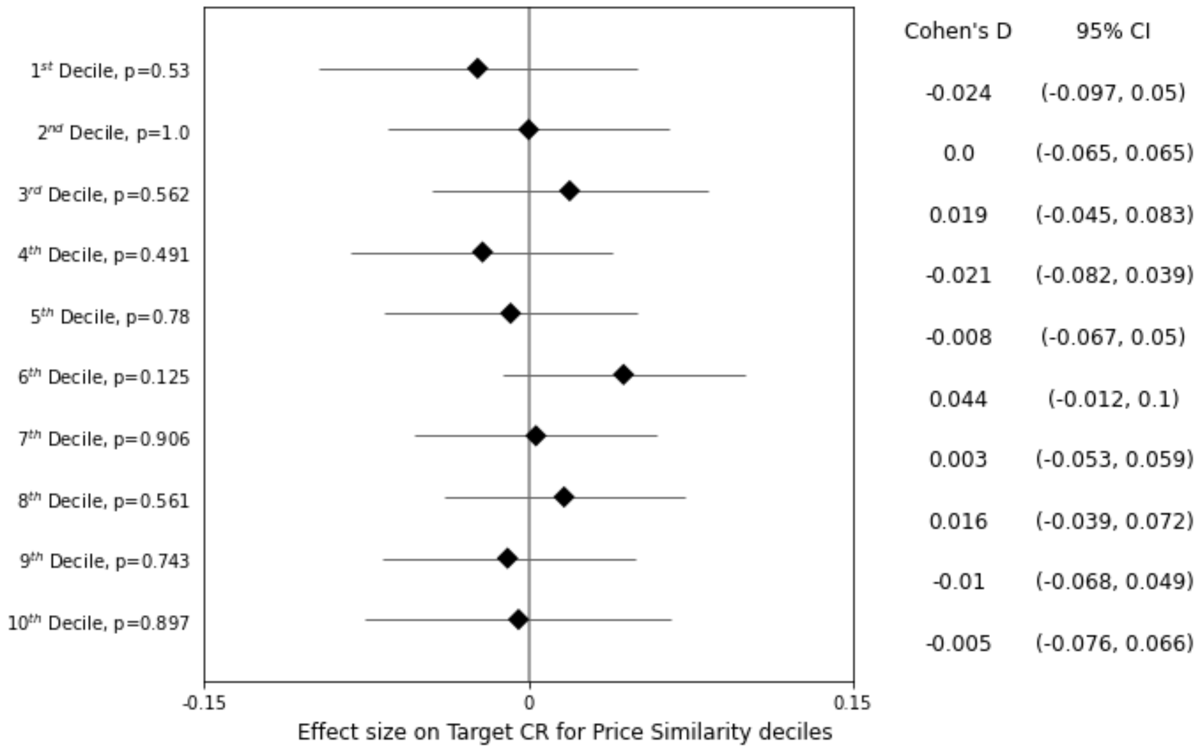
B-4) Price similarity



(a)



(b)

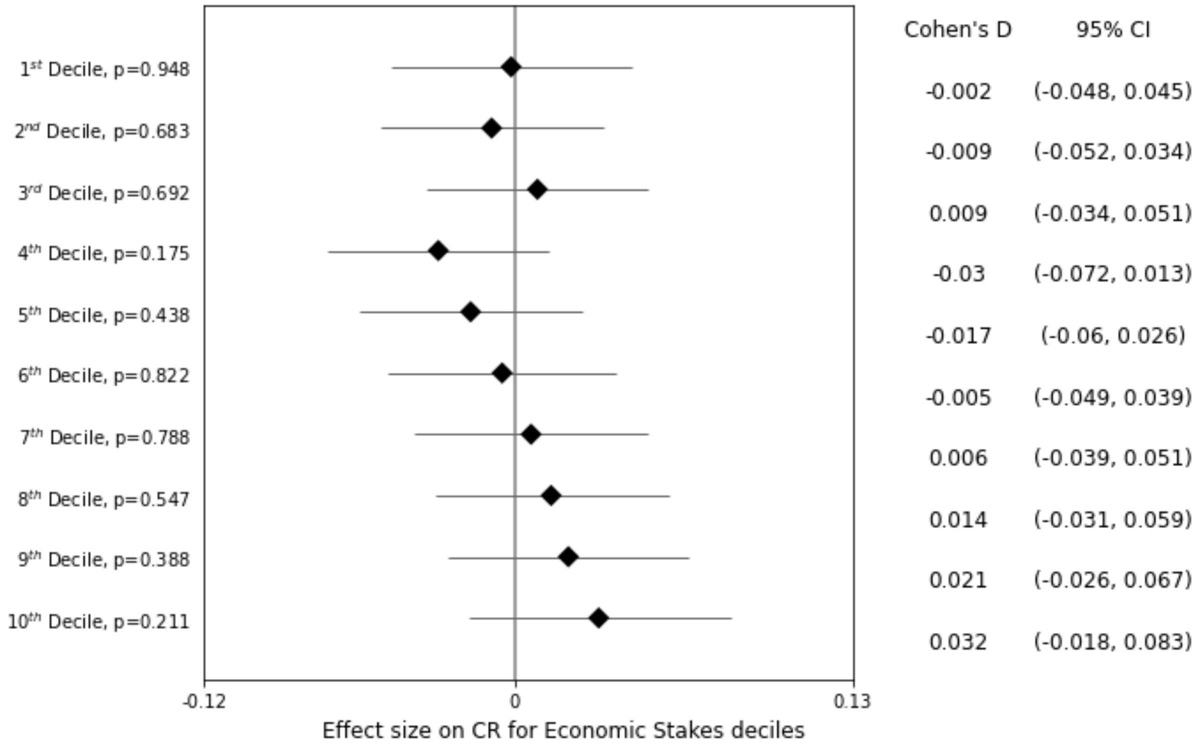


(c)

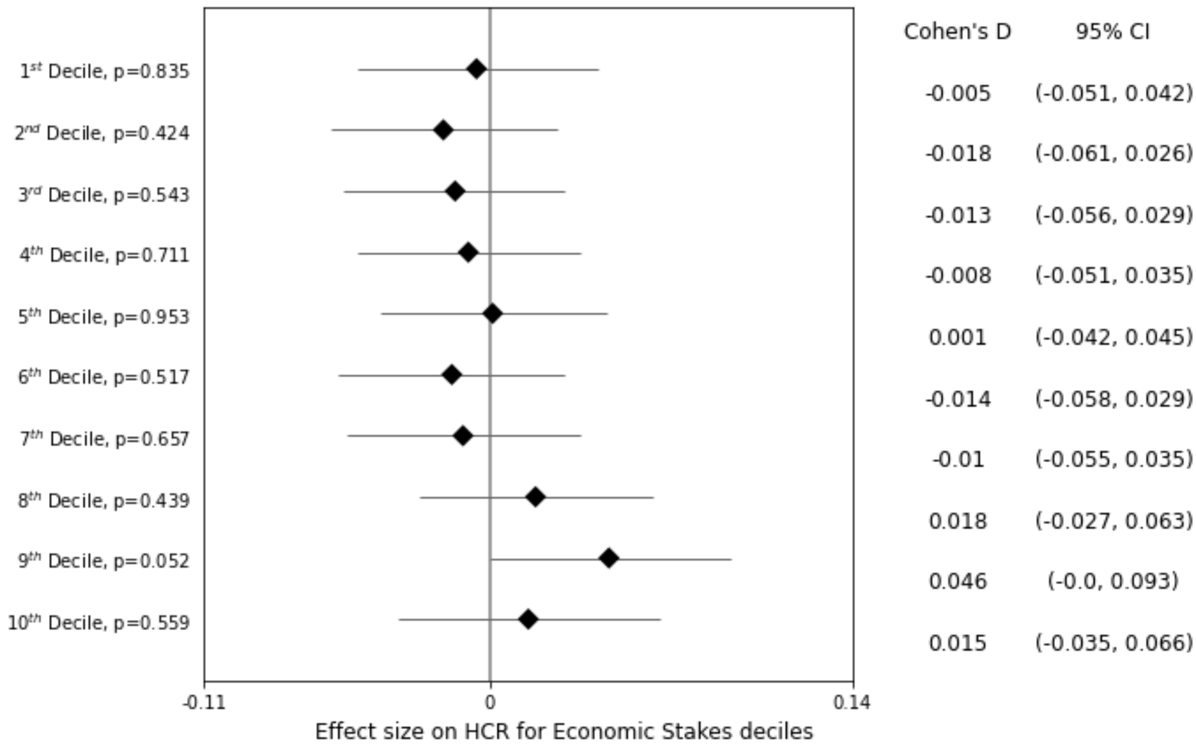
Figure B4 : Effect size estimated for each decile of price similarity.

Note: Figures show respectively the effect size for (a) the overall conversion rate (CR), (b) the highlighted conversion rate (HCR), and (c) the difference between the target and the competitor conversion rate. Each line represents the effect size (Cohen's d), and 95% confidence interval estimated based on subsamples of the sessions that correspond to each decile in terms of price similarity. The price similarity of a search session is defined as the ratio between the price of the Fastest and the price of the Cheapest itineraries $\left(\frac{P_F}{P_C}\right)$*. The lower the ratio, the higher the similarity (1st decile:* $\frac{P_C}{P_F} \geq 3.18$ *; 2nd decile = [2.31, 3.18[; 3rd decile = [1.91,2.31[; 4th decile[1.66, 1.91[; 5th decile = [1.48, 1.66[; 6th decile = [1.35, 1.48[; 7th decile = [1.25, 1.35[; 8th decile = [1.16, 1.25[; 9th decile = [1.08, 1.16[; 10th decile = [1, 1.08[ ).*

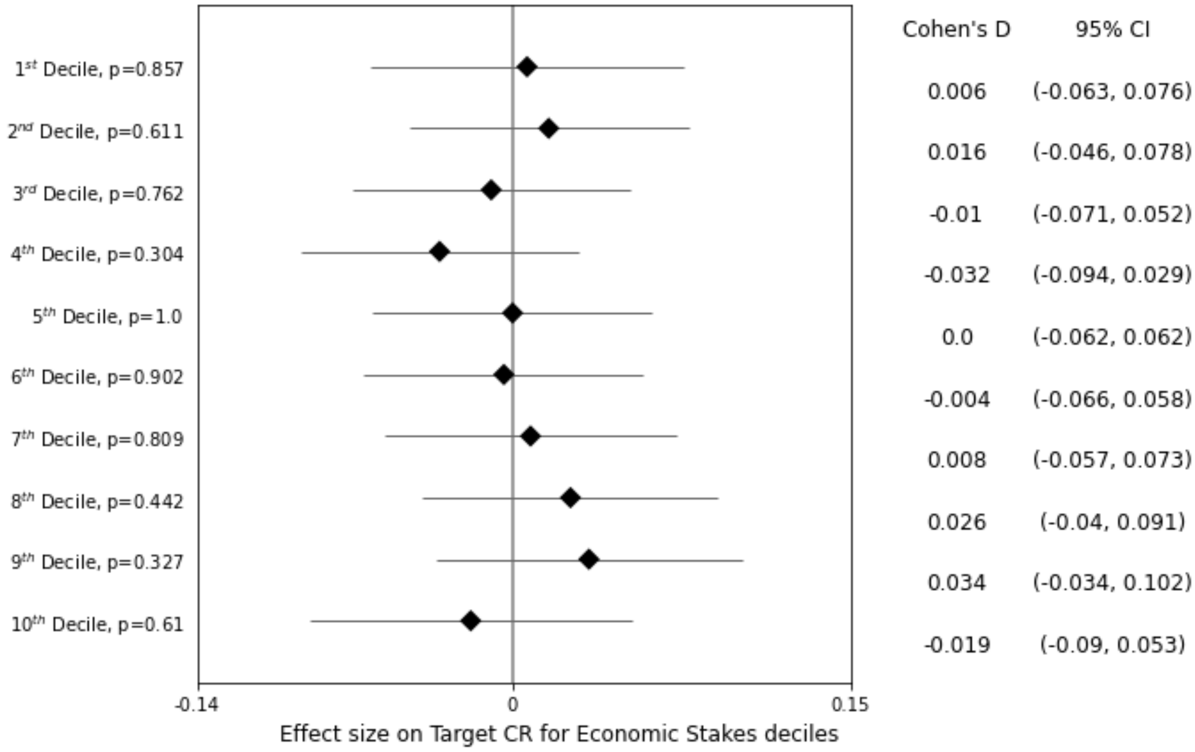
B-5) Economic stakes



(a)



(b)

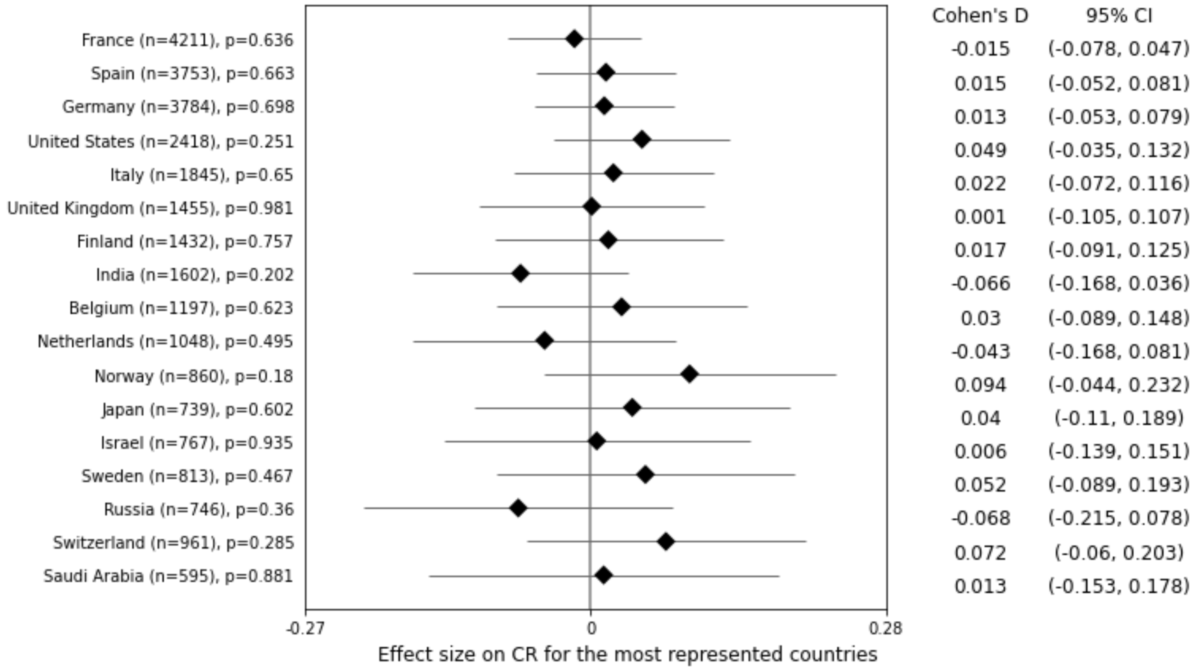


(c)

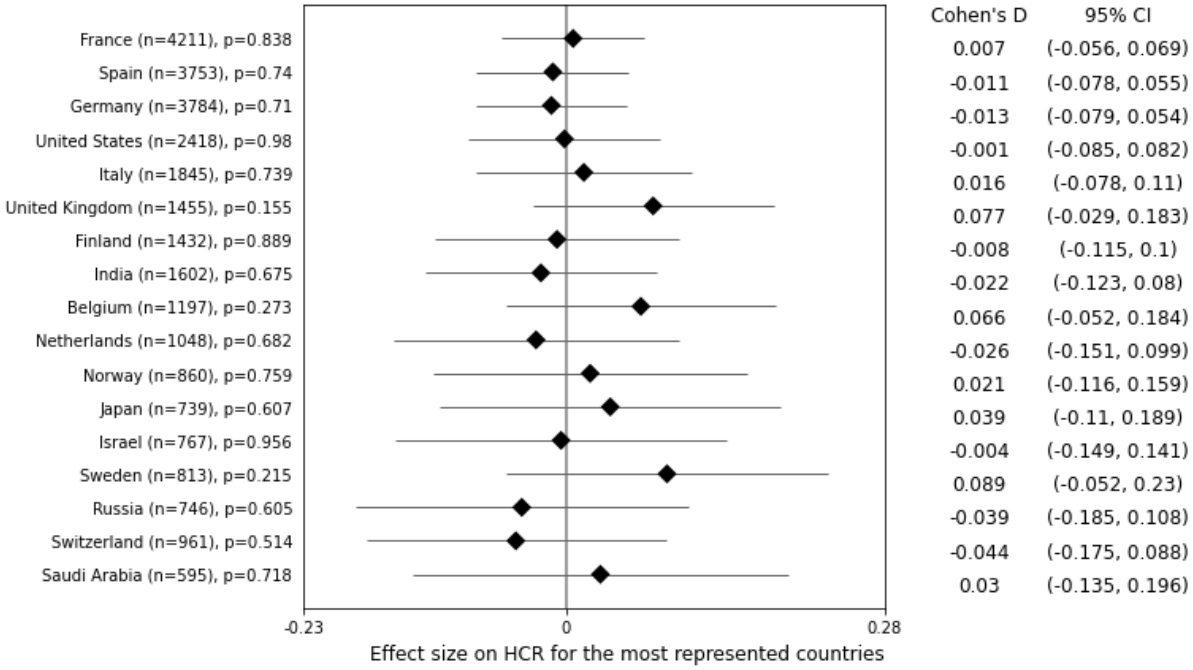
Figure B5: Effect size estimated for each decile of economic stakes.

Note: Figures show respectively the effect size for (a) the overall conversion rate (CR), (b) the highlighted conversion rate (HCR), and (c) the difference between the target and the competitor conversion rate. Each line represents the effect size (Cohen's d), and 95% confidence interval estimated based on subsamples of the sessions that correspond to each decile in terms of their economic stakes. The economic stakes of a search session is defined by the price (in USD) of the Cheapest itinerary (P_c). The higher the price of the Cheapest, the higher the economic stakes (1st decile: $P_C \in$ [0, 112[; 2nd decile = [112,167[; 3rd decile = [167,224[; 4th decile[224,295[; 5th decile = [295,388[; 6th decile = [388,510[; 7th decile = [510,643[; 8th decile = [643,826[; 9th decile = [826,1208[; 10th decile: [1208, $+\infty$[ ).

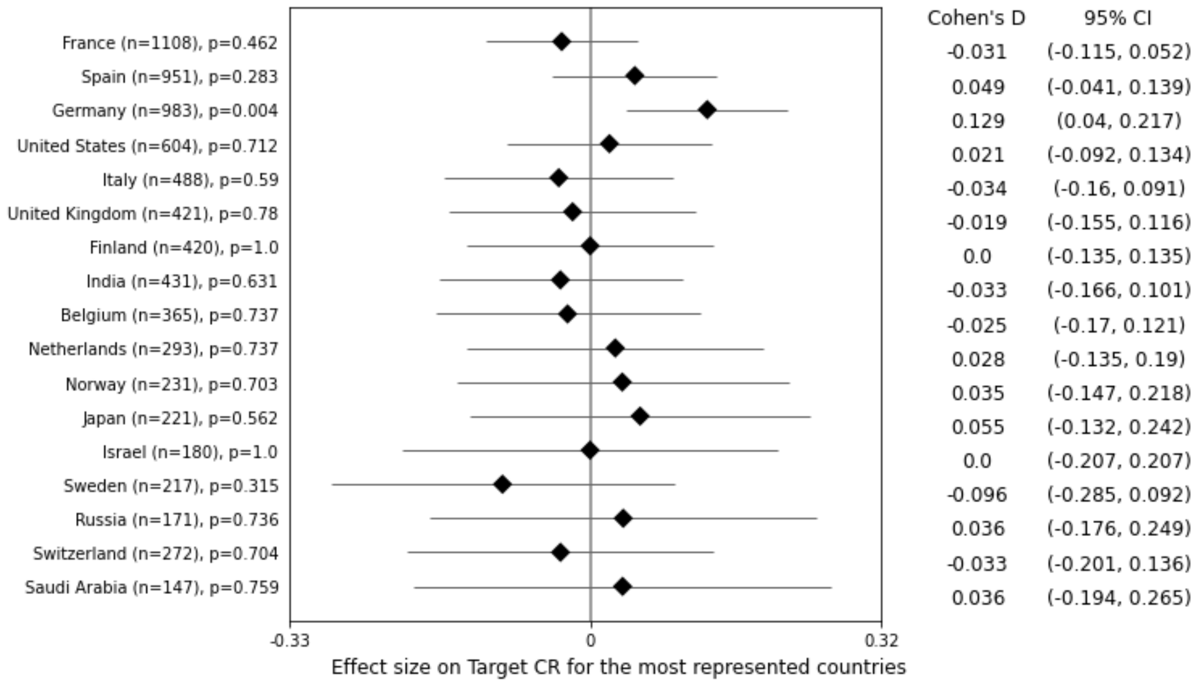
B-6) Countries



(a)



(b)



(c)

Figure B6: Effect size estimated for the countries where most of the sessions took place.
Note: Figures show respectively the effect size for the (a) overall conversion rate, (b) highlighted conversion rate, and (c) difference between targeted and competitor conversion rate. Each line represents the effect size (Cohen's d), and 95% confidence interval estimated based on subsamples of the sessions that correspond to the countries where most of the searches were made. For each country, n indicates the number of users from that country.