

Information-seeking strategies in medicine queries: a clinical eye-tracking study with gaze-cued retrospective think-aloud protocol

Tim Muntinga, University of Oxford, Radcliffe Observatory Quarter, Woodstock Road, Oxford. OX2 6GG. Tel: +44 1865 289300, tim.muntinga@phc.ox.ac.uk

Greg Taylor, University of Oxford, 1 St Giles, Oxford OX1 3JS. Tel: +44 1865 287210, greg.taylor@oii.ox.ac.uk

ABSTRACT

BACKGROUND: Medicines are increasingly purchased online, yet little is known regarding the ocular information-seeking behavior with medicine queries in search engines. A share of pharmacies found via search engines operate unlicensed and sell prescription-only medicines without a prescription. This study aimed to investigate how search engine users distinguish between genuine and falsified sources of information in terms of unlicensed and licensed online pharmacies in the case of medicine queries. **METHODS:** Eye-tracking of search tasks (transactional, navigational, informational and two limited results) in a Google search engine environment with retrospective gaze-cued think aloud protocol. Purposive sample of N=50 across three hospitals and one general practitioner. **RESULTS:** Discovery of a trichotomy of ocular search strategies based on the inclusion or exclusion of URLs in the information-seeking process. Finding of dissonance to existing studies related to fixation duration of search engine result page (SERP) elements. Discovery of an addition to information foraging theory: proximal cues are, in environments with non-credible information, used in both positive and negative ways.

Keywords

Eye-tracking, medicine queries, search engines, medical information-seeking behaviour, mixed methods.

1. INTRODUCTION

Health-related information found online is of varying quality (Eysenbach, Powell, Kuss, & Sa, 2002; Ilic, Risbridger, & Green, 2004; Jansen & Spink, 2006) and is usually retrieved via search engines (Spink et al., 2004; White & Horvitz, 2009). A significant amount of such health-related queries are related to online pharmacies with globally approximately 520,000 queries submitted to Google every day (Google Adwords Keyword Planner, 2014; Markey, 2007; White, Dumais, & Teevan, 2009). However, significant amounts of websites pretending to be pharmacies operate unlicensed, mitigating medical supervision by selling prescription-only medicines without a prescription. Medicines purchased from these sources have been regularly found to be of poor quality, lacking or containing the incorrect active pharmaceutical ingredient and missing patient information leaflets.

Various reports suggest that patients face considerable difficulty distinguishing unlicensed pharmacies from genuine ones. For instance, Solomon (2007) and Billerbeck (2011) report lethal metallic intoxication following medicine bought online: "It looked like a reputable site. [...] The box of pills that he [the coroner] had found in her home, were the same, as you would get at a pharmacy. Everything looked legitimate". Similarly, a dummy pharmacy set-up in 2012 by a Dutch Ministry achieved a top three organic position for keywords such as "sleeping pills" and "antidepressants" (mostly prescription-only medicines). If the payment page had not been disabled, 161,500 prescription-only medicines without a prescription would have been sold (Rijksoverheid, 2012).

Pharmacies requiring no prescription have been present via sponsored and organic search results. Search engines, such as Google, have now taken (legal) action and removed such illegal pharmacy advertisements. However, more effort is needed to protect organic search. An interminable cat-and-mouse game evolved between search engine algorithms and unlicensed online pharmacies employing black-hat search engine optimization techniques to gain high presence. These unethical practices, if able to mitigate Google's algorithmic anti-spam measures, allow unlicensed pharmacies to attract credulous buyers. For instance, non-accredited pharmacies have been found prominent in search engines for shortage drugs, while accredited pharmacies are little visible (Liang & Mackey, 2013).

This paper aims to understand the ocular strategies employed with querying behavior in medicine queries, and how search engine users can distinguish between licensed and unlicensed sellers of medicines. Despite clinical case reports highlighting

treatment failures and fatalities attributable to medicines obtained from such unlicensed outlets (Barber & Jacyna, 2011; Burger, Bosma, & van der Klooster, 2010; Levesque, 2004; Lineberry & Bostwick, 2004; Müller, Weinmann, & Hermanns-Clausen, 2009), little is known about the information-seeking behavior of patients regarding medicine queries in search engines.

Therefore, this paper has the following three objectives supporting the above stated aim:

1. To understand how participant characteristics and query types influence the ability to distinguish genuine from unlicensed online pharmacies.
2. To understand the role of search engine result page (SERP) credibility factors (title, URL, abstract and review stars) in distinguishing between genuine and unlicensed sellers of medicines.
3. To understand the role of proximal cues in information-seeking to distinguish between genuine and unlicensed sellers.

2. RELATED SEARCH ENGINE INFORMATION-SEEKING STUDIES

2.1 Participation characteristics

Generalizability of information-seeking behavior studies is debatable as the majority of (non-log analysis) papers have focused on (American) undergraduate college students (see Appendix 1). Despite such focus, information seeking behavior has been found to differ by continent (Jansen & Spink, 2005), gender (Lorigo et al., 2006; Roy & Chi, 2003) and age group (Bilal & Kirby, 2002). Student search engine users are reported to be "heavily influenced by the order in which the results are presented, and, to a lesser extent, the actual relevance of the abstracts" (Pan et al., 2007, p. 318) and predominantly view the first two organic results (Granka, 2004). Cutrell & Guan (2007) report that virtually homogenous attention is paid to each search engine result page (SERP) element (title, URL, abstract). However, Eysenbach & Köhler (2002, p. 576) found no participants conducting source evaluations. Advancing this dispute, Kammerer & Gerjets (2012) report incongruence to Eysenbach & Köhler (2002) discovering that students accessed more "objective information than [...] subjective or commercial information" (p. 94), thus employing source evaluations.

The low amount of yearly medicine prescriptions received by student population questions the generalizability of previous work towards online pharmacy selection behavior. In The Netherlands, ages employed in these studies accumulate between

3.3 to 4.6 prescriptions per year, compared to 12 in the age group 41-64, and 23.8 in the age group 65-69 (Stichting Farmaceutische Kengetallen, 2013, p. 18). Nevertheless, results from a previous study on student-based populations hint at the potential risk of selecting an unlicensed pharmacy. Administrating two webpages embedding untrustworthy features found in such pharmacies, Ivanitskaya et al. (2010) found 22% (total N=1914) of the students willing to recommend a rogue pharmacy to family and friends. Those that “used general search engines had worse evaluation skills than students who reported more traditional methods for making health decisions” (p. 58), hinting at potential variance by search engine usage. Evidence exists of differing information seeking behavior by query type (Broder, 2002; Granka, 2004; Rose & Levinson, 2004; Terai et al., 2008), and Granka (2004) found the majority of participants viewed the first two results. Therefore, we included multiple participant characteristics and query types, including a result page only showing the two first result snippets.

2.2 Credibility

Information-seeking choices are based on (perceived) credibility, a key criterion in decisions whether to accept or reject retrieved information (Hilligoss & Rieh, 2008). It is commonly agreed that trustworthiness and communicated expertise are sub factors of (perceived) credibility (Rieh, 2010; Robins, Holmes, & Stansbury, 2010). In consonance with Fogg & Tseng (1999) trustworthiness is in this paper taken to mean the perceived goodness or morality of the source, thus stressing the earlier mentioned source evaluation, while expertise refers to the source’s perceived knowledge or skill (e.g. a doctor versus a layperson providing medical advice). Querying for “purchase antibiotics without prescription” yields many unlicensed online pharmacies, employing black hat techniques to, besides increasing their SERP position, embed credibility in the form of fake ratings and falsified URLs (e.g. using 302 URL hijack exploits). A wide array of credibility typologies are defined concerning information (cf. Rieh, 2010, p. 1339), however, in search engine environments, a restricted amount of information is available for credibility assessment.

2.3 Information Foraging Theory

Information foraging theory (Pirolli, 2007) forms a theoretical basis for the selection of search engine results (Kammerer, Gerjets, & Lewandowski, 2010). It posits that humans use similar strategies in information acquisition behavior as animals do when foraging for food. The fundamental notion of information selection is seeking maximum benefit within the minimum amount of effort. Similar to animals that “gravitate towards the most filling food that requires the least amount of effort to consume” (Russell-Rose & Tate, 2012, p. 28), humans seek the strongest “information scent” which determines the likelihood of selection of a particular search result. The strength of this information scent is the perceived congruence of the user’s

information goal with the visible proximal cues. For example, medical information search involves proximal cues related to one's current informational goal, such as "doctor" or "medicine" that hint at the topical relevance of the source. A stronger information scent, suggesting greater relevance, increases the likelihood of selection of a result (Carroll, 2003).

Information foraging theory predicts a sequential method of interaction with SERP abstracts; advancing to the next result when all information is "consumed" or the termination of the information search when satisfied. The problem with unlicensed pharmacies is that these results often display alongside genuine pharmacies; hence, a sequential treatment of results would theoretically not yield most satisfactory results. Kammerer et al. (2010) highlight that, under information foraging theory, perceived information credibility, and information quality do not play a role in SERP selection; only SERP rank and information scent (proximal cues) matter. Information foraging theory does, besides this, not encompass a specific prediction of behavior related to falsified information sources.

3. METHODS

3.1 Overview of methods

Mixed-methods eye-tracking with retrospective gaze-cued think-aloud protocol and a short survey. Participants were administered five information-seeking search tasks in a simulated Google environment with the ability to view the landing page. All participants were presented with the same keywords and set of results to maintain consistency. The first three search tasks (navigational, transactional and informational) displayed a full first search engine result page (see Figure 1) with the ability to view the landing page of each linked site. The last two search tasks were simplified and contained a choice of two options, mimicking the fact that the majority of search engine users click on the first two results (Granka, 2004).

Figure 1. Example SERP of transactional query and example landing page

Landing pages and SERP results were, in all cases, constructed based on genuine and unlicensed pharmacies found in Google for popular pharmaceutical compound keywords (Google Adwords Keyword Planner, 2014). Queries executed included keywords "Cialis", "sildenafil citrate", "Viagra" and cardiovascular drug digoxin with brand name Lanoxin. Unlicensed status of

pharmacies was verified based on academic work describing characteristics of such outlets (Fittler, Bószé, & Botz, 2013; Gelatti et al., 2013; Ivanitskaya & Brookins-Fisher, 2010; Orizio, Merla, Schulz, & Gelatti, 2011), a pharmacy licensing verification database, and prescribing legislation. Landing pages and SERP results were slightly linguistically and visually modified to comply with ethical constraints. On the landing pages of actual online pharmacies and search engine results, pharmaceutical names were replaced by fictitious names for similar ethical reasons. We recorded the gaze behavior, mouse behavior, and screen behavior. This was followed by a compact survey delineating participant characteristics that included: gender, age, search engine experience ("How often did you use Google last week?"), knowledge of Boolean operators ("Do you know what Boolean operators are (e.g. AND, OR)"), buy amount online and whether the participant had medical expertise.

3.2 Gaze-cued retrospective think-aloud protocol

After finishing the search tasks and survey, participants were shown a screen recording of their whole session (see Figure 2) with an overlay of their eye-behavior, consisting of fixations (looking more than 100 ms at a specific point) and saccades (quick eye movement).

Figure 2. Example of participant-view in gRTA

gRTA (gaze-cued think aloud protocol) is a verbal protocol eliciting reasons and underlying cognitive processes by asking for the narration of conducted tasks (Ericsson & Simon, 1980). Presenting users with their gaze replay mitigates fabrication of verbal reasoning as the actual behavior is shown back simultaneously. Hence, it is a valid and reliable method in eye-tracking research (Guan, Lee, Cuddihy, & Ramey, 2006). Respondents were instructed to talk about the reasons for looking at certain points, in particular SERP page elements (URL, title, abstract and credibility cues) and were probed if silent. Retrospective think aloud was chosen instead of concurrent think aloud as it mitigates negative influence that the cognitive burden of thinking aloud has on task performance; thus enhancing ecological validity (Van Den Haak, De Jong, & Jan Schellens, 2003). gRTA started within 30 seconds of finishing the eye-tracking. Furthermore, gRTA allows to acquire "thick description" (Geertz, 1994), understanding comments provided to the context of gaze behavior and triangulation (Denzin, 1970), which are two important factors for establishing credible qualitative research (Merriam, 2014; Tracy, 2010).

Search tasks started with the sentence "You heard from a relative that [fictitious medicine name] works for a problem you have. You will see a search engine result page. Your task: [information need instruction]". To enhance consistency, verbal protocols were standardized (Ericsson & Simon, 1980). This included protocols for participant recruitment, explanation of the study,

(re)calibration and various (predicted) participant mistakes (e.g. forgetting information need instruction). gRTA probes were standardized as well.

3.3 Sampling

A total of 64 patients were recruited using non-probability purposive sampling (inclusion criteria: age groups from 18 to older than 60). The appropriate sample size was constructed inductively based on theoretical saturation of results by age. Twelve subjects were excluded whose cornea could not be accurately calibrated by the eye-tracker. Two participants were found satisficing (i.e. clicking without fixations) and were also excluded from the analysis. The total purposive sample of N=50 consisted out of n=18 participants below 40 years of age, n=32 participants above 40 years of age; n=13 with medical expertise; n=18 males and n=32 females; n=1 indicated no Google use, n=20 indicated less than 25 times of Google use last week, n=29 indicated more than 25 times last week. Patients were recruited from three hospitals and one general practitioner's office across the Netherlands, covering three different areas (South, West and North). The portable eye-tracker was set-up in a separate room with consistent light conditions. In all cases, theoretical saturation occurred after 4-7 similar respondent characteristics. The sample size is similar to or larger than that used in similar eye-tracking research (see Appendix A).

3.4 Eye-tracking

We used a Tobii-X2 60Hz eye-tracker mounted on a 14" laptop (1366x768 pixels resolution). We used a nine-point calibration. Illumination was consistent around 600 lux. Data was recorded and partially analyzed in Tobii Studio Enterprise. The chronological threshold for fixations (i.e. the minimum amount of milliseconds for which a participant must look at a point for a fixation to be recorded) was set at 100 ms.

3.5 Qualitative analysis

A modified analytic induction strategy based on a directed content analysis approach guided coding development. After that, descriptive codes were cyclically developed in vivo based on gRTA. Finally, pattern matching and a between-group contrasting strategy was used to discover links and to triangulate. Coding was done directly after gRTA to support theoretical saturation findings as well as later in totality with all N=50 gRTA sessions. After each gRTA session, quantitative data was used to validate qualitative coding. For instance, a participant's claim to have considered only URLs could be readily cross-checked against their actual eye behavior. This verification helps to guard against the threat of fabrication and participant satisficing.

3.6 Ethics

The University of Oxford provided ethical approval for this project, which included informed consent and full debriefing.

4. RESULTS

4.1 Association with participant characteristics

Table I summarizes the strength and direction of association of unlicensed pharmacy selection. This measure of association is denoted by γ and ranges from -1 (strong negative association), through 0 (no association), to 1 (strong positive association). Thus, a $\gamma = 0.9$ implies that people with the characteristic in question have a (proportionally) strong propensity to select unlicensed (rogue) pharmacies, whereas a $\gamma = -0.2$ would suggest that the characteristic is weakly associated with a tendency to select genuine pharmacies. This table shows that unlicensed pharmacy selection varies by query type and participant characteristics.

Table I. Summary Goodman Kruskal’s Gamma by participant characteristic with propensity to select an unlicensed pharmacy

Qualitative analysis of all participants’ gaze paths revealed that the susceptibility to unlicensed, rogue pharmacy selection is related to search strategies and differs by query type. One respondent summarized the search strategy variation by query type found in many participants:

“I never look and click at the first result. You are able to purchase this position by smart webpage programming. But, I have to say, when I query for something such as ‘when did Columbus die’, getting Wikipedia-like things, I click the first one directly. When I seek very factually, I take the first one. When buying, I never do this.”^a

^a = (Participant chose a genuine pharmacy in the navigational query, rogue in informational query, was older than 40 and reported more than 25 weekly Google queries).

4.2 Fixation duration

Deviation of search behavior by query type was also seen in a comparison of the number of fixations on each search element in comparison with the type of pharmacy selection (see Table II). Participants who selected a genuine pharmacy on the navigational search task looked at page titles an average of 17.61 times, whereas those who chose a rogue pharmacy did so

only 8.63 times. Throughout query types, participants selecting genuine pharmacies have more fixations across all elements. In particular, those selecting a genuine pharmacy fixated more on the URL (mean = 19.29, SD = 16.45) than those selecting a rogue pharmacy (mean = 5.06, SD = 6.84). A Wilcoxon Mann-Whitney test is used to understand whether fixation count across SERP elements differs statistically significantly between those selecting genuine and rogue pharmacies.

Table 2. Differences of number of fixations on SERP elements by pharmacy selection.

4.3 Navigational query

A majority of participants selecting a rogue pharmacy in the navigational task neglected to read abstract URLs and were, therefore, susceptible to such mistakes. Furthermore, mean fixation duration on URLs is found to be longer for those who selected a genuine result; despite participants selecting rogue pharmacies exhibiting a wider spread (see Figure 3).

Figure 3. Comparison of fixation count (left) and fixation duration (right).

Younger participants (< 40 years) and those knowledgeable regarding Boolean operators—a proxy for search engine expertise—often mention during gRTA that they scout for a “clean” [Dutch: schone], “simple” [Dutch: strak] or “bare” [Dutch: kale] URL:

“I find such a long URL strange [points at first result]. That rating is strange as well [points at second result]. I never consciously thought that the first result would be okay. Trustworthy is clean and simple.”^a

^a = (Participant was younger than 40, reported more than 25 weekly Google queries and chose consistently genuine pharmacies across all query types).

Gaze paths in conjunction with the number of fixations per SERP element (cf. Table II) showcase two distinct patterns. First, participants above 40 years old mostly employ a title-based strategy, failing to recognise a rogue pharmacy, whereas those below 40 years old utilise URLs, therefore consistently selecting genuine pharmacies. Figure 4 exhibits two typical fixation overlays with corresponding fixation duration that showcases this separation. The left fixation overlay presents a young (age <

40) user with more than 25 weekly search queries. The right shows a typical older (age > 40) participant with less than 25 weekly search queries.

Figure 4. Fixation duration comparison

Circles in the fixation overlay denote a fixation (looking more than 100 ms at a specific point, e.g. reading) with size hinting at fixation duration. Saccades (quick eye-movements), not cognitively processed, are presented with a line between fixations. The left image displays an efficient strategy of various fixations on the URL, selecting a genuine pharmacy, contrasted by the right employing a scattered ocular strategy at titles and descriptions, selecting a rogue pharmacy.

During the gRTA phase, those employing a title-based search strategy alluded to titles and descriptions as significant sources of proximal cues. These users also consistently selected rogue pharmacies. Participants mentioned looking for “kring pharmacy” (name of the pharmacy used in search task) and “pharmacy” and/or “online” as a word, despite the fact that these SERP elements are easy to fabricate:

“I merely looked for the word ‘kring’. I chose the first one as I sought a Kring Pharmacy. I did not look further, as I directly found what I was looking for.”^a

^a = (Participant was older than 40, reported less than 25 weekly Google queries and chose consistently rogue pharmacies across all query types).

Exceptions to this rule are various sceptical participants employing “negative” proximal cues in their search. Abstracts displaying such perceived negative proximal cues are not pursued further. Examples often mentioned include “100% trustworthy” and “trustworthy and cheap”. This search strategy resulted, in all cases, in the selection of a genuine pharmacy and is visible across all query types:

“Firstly, I looked at all the URLs. Then, I looked at those that really want to be trustworthy. Those that mention it the most, that trustworthiness [Dutch: betrouwbaar] are often the least. I don’t believe that.”^a

^a = (Participant was younger than 40, reported more than 25 weekly Google queries and chose a genuine pharmacy in this navigational query).

Two aggregate heatmaps of multiple users reveal deviation of fixation patterns with the varying outcome (Figure 5). The left image portrays an aggregation of gaze behavior of users selecting a rogue pharmacy. Perceptible are little fixations on

“falsified” top level domains (e.g. .com.cn and .in). The right image shows fixation patterns of participants selecting a genuine pharmacy and significant fixations on URLs and top level domains.

Figure 5. Left: Aggregate of fixations of multiple users selecting rogue pharmacies: unable to spot falsified top level domains of first two results (.com.cn and .in). Right: Aggregate of fixations of multiple users selecting genuine pharmacies: reading URLs and top level domains.

4.4 Transactional query

Search engine result pages that embed rogue pharmacies often include results that exhibit warning messages from governments, anti-counterfeit medicine organizations or blacklists. The first two results of this query task mentioned the availability of medication for purchase online without a prescription. The last three abstracts embedded such warning messages, mentioning that Aliraprovil is not available without a prescription. Two respondents did specifically mention not wanting to buy Aliraprovil online and resisted making a decision in the transactional task after reading the bottom three abstracts:

“[Researcher asks: what is the reason you did not want to purchase?] Because the third result mentioned that websites that sell Aliraprovil without prescription are mala fide sellers. So I do not buy it.”^a

^a = (Participant is a medical expert, younger than 40, more than 25 weekly Google queries, possession of knowledge of Boolean operators and chose consistently genuine pharmacies across all query types)

Congruent with other query types, the majority of participants selecting genuine pharmacies mentioned the use of negative proximal cues allowing them to filter out non-trustworthy results. Abstracts promoting themselves as “trustworthy” are counter effective:

“If the title mentions ‘trustworthy’, I find it less attractive. I think that you should not need to mention this. These others [points at various abstracts of unlicensed pharmacies] are too blatant. If it is about sound music speakers, it should be hip, not with medicines.”^a

^a = (Participant was older than 40, more than 25 weekly Google queries and chose consistently genuine pharmacies across all query types).

In sharp contrast, a few cases, consisting mostly of older respondents with less than 25 queries per week used these *negative* proximal cues as *positive* proximal cues in their selection behavior:

“I chose this one as it felt safe. The words ‘trustworthy’ and ‘pharmacy’ in the title convince me.”^a

^a = (Participant is a medical expert, older than 40, less than 25 weekly Google queries, bought more than once per month online and chose consistently genuine pharmacies across all query types).

4.5 Informational query

Search behavior in the informational query, which asked participants to determine whether Maqelaxhophan was available with or without a prescription, did not display an unambiguous pattern as exhibited in previously discussed queries. Participants using a URL-based strategy often included titles resulting in more genuine pharmacy selection ($p = 0.0038$ and $p = 0.0277$). Those selecting genuine pharmacies mentioned seeking a balance between positive (e.g. "hospital") and negative proximal cues (e.g. "risk", "side effects"), in contrast to a binary usage before, as summarized by a participant:

"I looked at particular words which I have a high degree of trust with, 'risk', 'complications' and 'hospital', I choose it because they also mention negative words and the complications."^a

^a = (Participant was older than 40, reported more than 25 weekly Google queries and chose consistently genuine pharmacies across all query types).

4.6 Limited choice query (two abstracts)

This task embedded two abstracts and saw similar outcomes to the navigational queries: a URL-based strategy was employed by nearly all participants who selected genuine pharmacies, including all participants with knowledge of Boolean operators. Typical behavior was a sequential process, in which a participant briefly examined titles first, then proceeded to URLs (see Figure 6 for a visual representation). A title-based strategy employing "positive" proximal cues resulted, in congruency with the navigational search task, in unlicensed pharmacy selection.

Figure 6. Example of titles first strategy until 2 seconds 438ms (left), not satisfied, moves on to URLs until 4 seconds (right image 2:436-04:957).

Frequently mentioned positive proximal cues were "kring apotheek" and "free delivery" in conjunction with looking at stars and textual ratings. Participants focusing on ratings chose more rogue pharmacies ($p = 0.0398$). Usage of perceived credibility factors and proximal cues, by various participants, is clearly visible:

"I did not look at the URLs; I was triggered by 'quick delivery' and 'free shipping'."^a

^a = (Participant was older than 40, reported more than 25 weekly Google queries and chose a rogue pharmacy in this limited query type).

Again, in congruence with navigational queries, users employing a URL-based strategy mention scouting for the “simplicity” of a URL to guide themselves:

“I chose the second result, here again: the URL is simple, while the first contains a dash. I never let stars or reviews guide myself.”^a

^a = (Participant was younger than 40, reported less than 25 weekly Google queries, had knowledge of Boolean operators and chose a rogue pharmacy in this limited query type).

5. DISCUSSION

5.1 Fixation duration discrepancy

Results display incongruence to several existing search engine studies: fixation duration measures across SERP elements significantly differ from each other, in sharp contrast to Cutrell & Guan (2007, p. 8) who reported homogeneity of fixation duration across SERP elements (title, URL, and snippet). This study embedded a wide variety of age groups (18-50, N=18). The present study, however, has discovered deviation in search strategy employed (e.g. URL versus title) by age, suggesting that different age groups should not be conflated. Furthermore, their study employed non-medical queries (e.g. finding the address of an airport).

5.2 Resolution to the source evaluation dispute

Contrasting evidence exists in search engine studies regarding the use of source evaluations (i.e. URL-based or trustworthiness of the website) in query behavior. Eysenbach & Köhler (2002) and Ivanitskaya, O’Boyle, & Casey (2006), found searchers not conducting source evaluations while the student-only sample of Kammerer & Gerjets (2012)’s suggested that students selecting more objective information do conduct source evaluations. Data from this study offers a resolution to this dispute. The results suggest that source evaluations are indeed conducted by younger participants, with more search experience and Boolean knowledge in similarity to the sample of Kammerer & Gerjets (2012). This group employs a URL-based source evaluation strategy and visits landing pages if required. This contrasts with reports of Ivanitskaya, O’Boyle, & Casey (2006). However, in congruency with Eysenbach & Köhler (2002), this study finds older participants (above 40) and with less search experience (< 25 weekly searches) not conducting source evaluations using URLs. It should be noted here that this group does visit landing pages, but either does not conduct source evaluations, or else does so ineffectively. Thus, the findings of this study can be viewed as reconciling the results of Kammerer & Gerjets (2012) with those of Eysenbach & Köhler (2002).

5.3 Participant characteristics

In this study unlicensed (rogue) pharmacy selection exhibited little variation by gender. However, the data shows a moderate association in the navigational query task ($\gamma = 0.51$). This is consistent with earlier work (Lorigo et al., 2006; Roy & Chi, 2003), which also found a significant role for gender that differed by query type. An important difference, though, is that Lorigo et al. (2006) and Roy & Chi (2003) found gender influencing *evaluation and information behavior*, while the present study discovered gender to be associated with the *outcome* of such behavior. gRTA and debriefing were used to gauge for any occurring learning effects; all users selecting a rogue pharmacy were found to mention not seeing it being an unlicensed pharmacy.

Existing search engine studies mostly used (American) college students as a population (cf. Appendix 1). Contrasting results of this study to existing literature could be explained by the use of a heterogeneous group (e.g. different age groups), use of medical queries, and variation in query types. Findings derived from the broader population embedded in this study reconcile, and extend results from Bilal & Kirby (2002) who found that age influences information seeking strategies. Specifically, this paper extends information foraging theory by discovering a trichotomous typology of search engine information seeking strategies differing by age (see Figure 7).

Figure 7. Three distinct information search strategies

Older users (> 40) mostly employ a “non-URL based strategy” (titles and descriptions), while younger users (<40) employ either an “exclusive URL-based strategy” or a “sequential title strategy” with verification by URLs. Such behavior differs per query type, which, in turn, corroborates the findings of other existing studies (Broder, 2002; Granka, 2004; Terai et al., 2008). Being not a true experiment—which is in information-retrieval eye-tracking studies often not intended—claims in this study are limited to, and based on association and correlation, rather than causation. However, ethical constraints resulted in the need to create a hypothetical set-up instead of real observed behavior in an actual SERP. Despite attempts for a verbatim resemblance of SERPs and landing pages, such constraints resulted in the need for several linguistic, visual and pharmaceutical modifications. Both genuine and rogue results were copied as close to reality as possible, yet these modifications might have impacted external generalization towards actual search results. However, several actions were

taken to limit this. First, SERP abstracts and landing pages were copied from multiple actual Google queries. Second, landing pages and SERP abstracts are adapted from real pharmacies (including verified rogue pharmacies listed on a blacklist). Third, existing literature offered characteristics of frequently used unethical techniques that drove selection of specific results and ethical modifications. Lastly, the SERP ranks were based on actual rankings: rogue pharmacies blossom due to black-hat techniques and reach higher positions than genuine pharmacies. Granka (2004) found the majority of participants clicked on the first few results, congruent to rogue pharmacy rankings. gRTA and visualized gaze paths show that users quite often looked beyond the first three results.

A second limitation is introduced by the ethical review board requirement of deploying fictitious medicine names to avoid participant deception. This has a bearing on the finding that medical education has no influence on selection behavior (e.g. $\gamma = -0.01$ for the navigational query). This finding suggests that, for instance, those with medical and non-medical expertise in this study showcased similar susceptibility to unlicensed pharmacies. If instead real pharmaceutical names were used, those with medical expertise would presumably be in a better position to use their topical knowledge. However, an appealing aspect of using fictitious drug names offsets this limitation. Unlike search tasks based on general knowledge, such as place names or historical facts, fictitious drug names eliminate unobserved heterogeneity in subjects' prior knowledge and therefore avoid spurious correlations induced by such heterogeneity. Lastly, Granka (2004) suggested that search behavior correlates with task difficulty. Every attempt was made to achieve a consistent difficulty across the various search tasks in this study. It might, however, be the case that comparisons across query types are affected by residual variations in task difficulty.

5.4 Impact on the design of online pharmacies and search engines

The findings of this paper provide a background to how the user interface design of licensed, online pharmacies in search engines could be modified. Vital for the design of solutions is to avoid entirely relying on the user's search literacy to distinguish licensed from unlicensed pharmacies. Embedding an image at licensed online pharmacies to display the licensing status can easily be forged and copied by the unlicensed counterparts and might therefore not yield the desired outcomes. The user interface of search engines could be modified in a way that excludes or annotated unlicensed sellers, or medical regulatory agency efforts should be conducted to build search literacy. Future studies could investigate the effectiveness of various SERP layout modifications aiding patients selecting online pharmacies. Furthermore, medical practitioners should be

aware of risks presented by online medicine searches and receive guidance to help patients in search of a genuine supplier of medicine.

6. CONCLUSION

The eye-tracking study with gaze-cued think-aloud protocol described in this paper discovered differing susceptibility to selecting unlicensed pharmacies by participant characteristics. Different ocular strategies based on looking at the URLs yielded different outcomes. Furthermore, based on qualitative comments, this paper argues that information foraging theory could be extended to situations with non-credible information as proximal cues used in both positive and negative ways. Future studies could for instance investigate how SERP user interface elements and proximal cues would require to be redesigned to contribute to a higher proportion of genuine pharmacy selection behavior in search engines.

7. SUMMARY OF ADDITION TO KNOWLEDGE

7.1 What did we know before this study?

- Students, even in health programs (medical education) are susceptible to unlicensed pharmacy selection.
- Search engine behavior mostly differs by age, gender and query type.
- Disagreement exists between papers describing source evaluations conducted in search engines.

7.2 What does this study add?

- Susceptibility differs by participant characteristic. Highest associations of susceptibility of unlicensed pharmacy selection: no Boolean knowledge, older than 40, less frequent Google user.
- Fixation count by SERP element differs by pharmacy selection and query type.
- Three distinct search strategies by SERP element: (1) URL-omission strategy: using titles and descriptions, (2) URL inclusion strategy, (a) exclusively URLs and (b) sequential URLs: titles in conjunction with URLs. First search strategy sees the highest amount of unlicensed pharmacy selection and is employed by participants older than 40 and with less search engine experience.
- Discovery of deviation by fixation duration of SERP element, in dissonance with Cutrell & Guan (2007).
- Proximal cues are, in environments with non-credible information, used in two ways: positive and negative.

8. ACKNOWLEDGMENTS

TM was funded during this study at the Oxford Internet Institute, University of Oxford, by the Koninklijke Hollandsche Maatschappij der Wetenschappen, Foundation De Korintheers, Dr. Hendrik Muller's Vaderlandsch Fonds, Prins Bernhard Cultuurfonds, VSB Fonds and KAS Bank. We are very grateful to Acuity-ETS and Tobii for lending the eye-tracking device and to the Oxford Internet Institute for providing the support during this study.

9. REFERENCES

- Barber, T., & Jacyna, M. (2011). Acute lead intoxication from medications purchased online presenting with recurrent abdominal pain and encephalopathy. *Journal of the Royal Society of Medicine*, 104(3), 120-123.
- Bilal, D., & Kirby, J. (2002). Differences and similarities in information seeking: children and adults as Web users. *Information Processing & Management*, 38(5), 649-670.
- Billerbeck, G. (2011). Glenda on rogue sites. YouTube (Fight Online Theft account). Retrieved from <http://www.youtube.com/watch?v=urGeA-d9Utw-t=133>.
- Broder, A. (2002). *A taxonomy of web search*. Paper presented at the ACM Sigir forum.
- Burger, M., Bosma, M., & van der Klooster, J. (2010). [Suicide attempt with antidepressants ordered from the Internet]. *Nederlands tijdschrift voor geneeskunde*, 155(48), A3661-A3661.
- Carroll, J. M. (2003). *HCI models, theories, and frameworks: Toward a multidisciplinary science*: Morgan Kaufmann.
- Cutrell, E., & Guan, Z. (2007). *What are you looking for?: an eye-tracking study of information usage in web search*. Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems.
- Denzin, N. K. (1970). *The research act: A theoretical introduction to sociological methods*: Transaction publishers.
- Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological review*, 87(3), 215.
- Eysenbach, G., & Köhler, C. (2002). How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. *Bmj*, 324(7337), 573-577.
- Eysenbach, G., Powell, J., Kuss, O., & Sa, E.-R. (2002). Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. *Jama*, 287(20), 2691-2700.
- Fittler, A., Bösze, G., & Botz, L. (2013). Evaluating aspects of online medication safety in long-term follow-up of 136 Internet pharmacies: illegal rogue online pharmacies flourish and are long-lived. *Journal of medical Internet research*, 15(9).
- Fogg, B., & Tseng, H. (1999). *The elements of computer credibility*. Paper presented at the Proceedings of the SIGCHI conference on Human Factors in Computing Systems.
- Geertz, C. (1994). Thick description: Toward an interpretive theory of culture. *Readings in the philosophy of social science*, 213-231.

- Gelatti, U., Pedrazzani, R., Marcantoni, C., Mascaretti, S., Repice, C., Filippucci, L., . . . Feretti, D. (2013). 'You've got m@ il: Fluoxetine coming soon!': Accessibility and quality of a prescription drug sold on the web. *International Journal of Drug Policy*, 24(5), 392-401.
- Google Adwords Keyword Planner. (2014). Google Adwords Traffic Estimator. Retrieved from <https://adwords.google.com/ko/KeywordPlanner>
- Granka, L. A. (2004). *Eye-R: Eye-Tracking analysis of user Behavior in online Search*: Cornell University, May.
- Guan, Z., Lee, S., Cuddihy, E., & Ramey, J. (2006). *The validity of the stimulated retrospective think-aloud method as measured by eye tracking*. Paper presented at the Proceedings of the SIGCHI conference on Human Factors in computing systems.
- Hilligoss, B., & Rieh, S. Y. (2008). Developing a unifying framework of credibility assessment: Construct, heuristics, and interaction in context. *Information Processing & Management*, 44(4), 1467-1484.
- Ilic, D., Risbridger, G., & Green, S. (2004). Searching the Internet for information on prostate cancer screening: an assessment of quality. *Urology*, 64(1), 112-116.
- Ivanitskaya, L., & Brookins-Fisher, J. (2010). Dirt cheap and without prescription: how susceptible are young US consumers to purchasing drugs from rogue internet pharmacies? *Journal of medical Internet research*, 12(2).
- Ivanitskaya, L., O'Boyle, I., & Casey, A. M. (2006). Health information literacy and competencies of information age students: results from the interactive online Research Readiness Self-Assessment (RRSA). *Journal of medical Internet research*, 8(2).
- Jansen, B. J., & Spink, A. (2005). An analysis of web searching by European AlltheWeb. com users. *Information Processing & Management*, 41(2), 361-381.
- Jansen, B. J., & Spink, A. (2006). How are we searching the World Wide Web? A comparison of nine search engine transaction logs. *Information Processing & Management*, 42(1), 248-263.
- Kammerer, Y., & Gerjets, P. (2012). Effects of search interface and Internet-specific epistemic beliefs on source evaluations during Web search for medical information: An eye-tracking study. *Behaviour & Information Technology*, 31(1), 83-97.
- Kammerer, Y., Gerjets, P., & Lewandowski, D. (2010). How search engine users evaluate and select Web search results: The impact of the search engine interface on credibility assessments. *Web search engine research*, 251-279.
- Levesque, C. A. (2004). *Tardive dyskinesia associated with Internet drug purchase*. Paper presented at the Mayo Clinic Proceedings.

- Liang, B. A., & Mackey, T. K. (2013). Online availability and safety of drugs in shortage: a descriptive study of internet vendor characteristics. *Survey of Anesthesiology*, 57(1), 54-55.
- Lineberry, T. W., & Bostwick, J. M. (2004). *Taking the physician out of "physician shopping": a case series of clinical problems associated with Internet purchases of medication*. Paper presented at the Mayo Clinic Proceedings.
- Lorigo, L., Pan, B., Hembrooke, H., Joachims, T., Granka, L., & Gay, G. (2006). The influence of task and gender on search and evaluation behavior using Google. *Information Processing & Management*, 42(4), 1123-1131.
- Markey, K. (2007). Twenty-five years of end-user searching, Part 1: Research findings. *Journal of the American Society for Information Science and Technology*, 58(8), 1071-1081.
- Merriam, S. B. (2014). *Qualitative research: A guide to design and implementation*: John Wiley & Sons.
- Müller, D., Weinmann, W., & Hermanns-Clausen, M. (2009). Chinese slimming capsules containing sibutramine sold over the Internet: a case series.
- Orizio, G., Merla, A., Schulz, P. J., & Gelatti, U. (2011). Quality of online pharmacies and websites selling prescription drugs: a systematic review. *Journal of Medical Internet Research*, 13(3).
- 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 Communication*, 12(3), 801-823.
- Pirolli, P. (2007). *Information foraging theory: Adaptive interaction with information*: Oxford University Press.
- Rieh, S. Y. (2010). *Encyclopedia of Library and Information Sciences*: Taylor & Francis.
- Rijksoverheid. (2012). *Factsheet internetpillen [Factsheet internetpills]*. Retrieved from <http://www.rijksoverheid.nl/bestanden/documenten-enpublicaties/brochures/2012/02/06/factsheet-internetpillen/factsheetinternetpillen.pdf>
- Robins, D., Holmes, J., & Stansbury, M. (2010). Consumer health information on the Web: The relationship of visual design and perceptions of credibility. *Journal of the American Society for Information Science and Technology*, 61(1), 13-29.
- Rose, D. E., & Levinson, D. (2004). *Understanding user goals in web search*. Paper presented at the Proceedings of the 13th international conference on World Wide Web.
- Roy, M., & Chi, M. T. (2003). Gender differences in patterns of searching the web. *Journal of Educational Computing Research*, 29(3), 335-348.
- Russell-Rose, T., & Tate, T. (2012). *Designing the search experience: The information architecture of discovery*: Newnes.

- Solomon, S. (2007). BC woman killed by fake drugs bought online: "Metal toxicity" from counterfeit pills reinforces danger of internet meds. . Retrieved from http://www.nationalreviewofmedicine.com/issue/2007/07_30/4_policy_politics_13.html
- Spink, A., Yang, Y., Jansen, J., Nykanen, P., Lorence, D. P., Ozmutlu, S., & Ozmutlu, H. C. (2004). A study of medical and health queries to web search engines. *Health Information & Libraries Journal*, 21(1), 44-51.
- Stichting Farmaceutische Kengetallen. (2013). *Data en feiten 2013: Het jaar 2012 in cijfers [Facts and figures 2013: The year 2012 in numbers]*. Retrieved from
- Terai, H., Saito, H., Egusa, Y., Takaku, M., Miwa, M., & Kando, N. (2008). *Differences between informational and transactional tasks in information seeking on the web*. Paper presented at the Proceedings of the second international symposium on Information interaction in context.
- Tracy, S. J. (2010). Qualitative quality: Eight "big-tent" criteria for excellent qualitative research. *Qualitative inquiry*, 16(10), 837-851.
- Van Den Haak, M., De Jong, M., & Jan Schellens, P. (2003). Retrospective vs. concurrent think-aloud protocols: testing the usability of an online library catalogue. *Behaviour & Information Technology*, 22(5), 339-351.
- White, R. W., Dumais, S. T., & Teevan, J. (2009). *Characterizing the influence of domain expertise on web search behavior*. Paper presented at the Proceedings of the Second ACM International Conference on Web Search and Data Mining.
- White, R. W., & Horvitz, E. (2009). Cyberchondria: studies of the escalation of medical concerns in web search. *ACM Transactions on Information Systems (TOIS)*, 27(4), 23.

APPENDIX 1: TABULAR OVERVIEW OF RELEVANT EXISTING SEARCH ENGINE STUDIES

| Study | Students | Subjects | Methodology | Main finding(s) related to this paper |
|------------------------------------|----------|-------------------------------------------------------------------------------------|--------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| (Cutrell & Guan, 2007) | No | 11 male and 7 female participants 18-50, diverse backgrounds (USA). | Eye-tracked search tasks | Fixation duration on URL, title and snippet is found to be equal, yet data is aggregated (no differentiation made by participant characteristics). |
| (Dumais, Buscher, & Cutrell, 2010) | No | 21 female and 17 male from a diversity of age groups. M age = 45.5 years, SD = 8.2. | | First three results are most looked at, congruent to Granka (2004) and Cutrell & Guan (2007). |
| (Granka, 2004) | Yes | 19 male and 15 female undergraduate students at a large university (USA). | | Search behavior differs by task type, gender and difficulty of the task. |
| (Jansen, Zhang, & Schultz, 2009) | Yes | 24 male and 8 female students of a major university (USA). | | Search engine brand (versus non-branded) has a substantial effect on evaluation (i.e. number of links examined). |
| (Kammerer & Gerjets, 2012a) | Yes | 10 male and 48 female undergraduate students at a German university. | | Contrasts findings of (Eysenbach & Köhler, 2002; Ivanitskaya et al., 2006) that searchers often omit source evaluations. Found that students have the skill to recognize higher quality information, selecting more objective than commercial or subjective abstracts. |
| (Aula, Majaranta, & Riih , 2005) | Yes | 11 female and 17 male students, M age 23.7 years. | | Different types of gaze-paths and evaluation styles of results by the user. |
| (Lorigo et al., 2006) | Yes | 14 male and 9 female undergraduate students at a large university (USA). | | Evaluation gaze paths differ by gender; success of information seeking differs by query type. |
| (Pan et al., 2007) | Yes | 11 male and 5 female undergraduate | | Little influence with ranks of abstracts swapped. Rank is more used as a selection-criterion than relevance. |

| | | | | |
|------------------------------------------------|-----|---------------------------------------------------------------------------------|-----------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------|
| | | students of Cornell University (USA). | | |
| (Gerjets, Kammerer, & Werner, 2011) | Yes | 10 male and 20 female students of a German university. | Eye-tracked search tasks with concurrent | CTA influences task outcome, influences ecological validity. Researchers recommend using RTA and refrain from CTA. |
| (Terai et al., 2008) | Yes | 5 male and 6 female Japanese undergraduate students. | think aloud protocol (CTA) | Scan path differs by query type, longer reading time in the informational query. |
| (Balatsoukas & Ruthven, 2012) | Yes | 9 male and 8 female undergraduate and postgraduate students at a UK university. | Eye-tracked search tasks with CTA and retrospective (RTA) | Different relevance judgment criteria affect fixation duration. Numbers of fixations on SERP elements (e.g. URL, title) differ by relevance criteria. |

|

Author biographies

Tim Muntinga

Tim Muntinga conducted this research as part of his MSc in Social Science of the Internet at the Oxford Internet Institute, University of Oxford. He has multiple years of consulting background in digital strategy, user experience research, and digital marketing, and he has a particular interest in the (digital) health intersection.

Greg Taylor

Greg Taylor is an economist whose research focuses on the economics of online markets and of markets for technology goods more generally. Special topics of interest include the search engine and online advertising industries, platform markets, consumer search behaviour and price comparison services, the attention economy, and online intermediary bias.