Online Search Stopping Behaviors: An Investigation of Query Abandonment and Task Stopping

Wan-Ching Wu
SILS
University of North Carolina
Chapel Hill, NC, 27599 USA
wanchinw@email.unc.edu

Diane Kelly
SILS
University of North Carolina
Chapel Hill, NC, 27599 USA
dianek@email.unc.edu

ABSTRACT
The goal of the study is to understand the factors that influence people’s search stopping behaviors during online information search. Past research on search stopping behavior has primarily focused on the stopping behavior that takes place at the conclusion of an information-seeking task. However, in this study we focus on two types of stopping behaviors that take place during information search tasks: query abandonment, or the point at which a person decides to stop his/her current query and enter a new one, and task stopping, or the point at which a person decides to stop the search task. A laboratory study was conducted with 48 participants who were asked to complete a set of six assigned search tasks and were interviewed about their experiences and search strategies after search. Results show that participants made query abandonment decisions based on the properties of search results, of queries and of search tasks. Their decisions to stop a task were influenced by the content they had examined, the goal they wished to achieve, the subjective perceptions they felt, and the study constraints they faced.

Keywords
Search Stopping behavior, Query Abandonment, Interactive IR, Search Strategies, Human-Information Behavior

INTRODUCTION
During information search, a person needs to make a series of decisions. Perhaps two of the most important decisions are determining when to issue a new query and when to stop gathering information for the search task. In the present study, we investigate people’s stopping behaviors as they interact with search engines to solve open-ended information problems. While previous work has highlighted the notion of “feeling good enough” as the main reason why individuals stop searching (e.g., Zach, 2005), most of this past research has focused on information seeking tasks, in which information gathering involves interactions with systems and human beings, and relied on recall from past search experiences rather than studying information search tasks, where actions are restricted to interactions with information search systems as they take place (Saracevic, 2010). Most of this past research has also only focused on the stopping behavior that occurs at the conclusion of a task, rather than the various stopping points that occur throughout the search, such as when a person decides to stop examining results for a query. In this paper, we focus on two types of stopping behaviors: query abandonment, or the point at which a person decides to stop his/her current query and enter a new one, and task stopping, or the point at which a person decides to stop the search task. Note that query abandonment and task stopping are different from search abandonment, which has been used in the context of large-scale search log analysis to describe instances when a person submits a query, but does not click on any of the search results (Joachims & Radlinski, 2007).

In this study, we seek to address the following research question: What factors contribute to people abandoning their queries and stopping search tasks? In our previous work (Wu, Kelly & Sud, 2014), we investigated query abandonment when participants were exposed to search engine result pages (SERPs) with different numbers of relevant results and different distributions of relevant results. Our findings showed that participants examined more results and clicked deeper when the first SERP contained more relevant search results. Participants also showed a tendency to abandon queries on the first SERP with a greater probability when there were fewer relevant results or when the results were displayed at lower ranks on the page. This paper extends our previous work by presenting qualitative results obtained from the same study, but which were not reported in Wu, Kelly and Sud (2014).
BACKGROUND

Query Abandonment

Query abandonment is defined as the point at which a person decides to stop his/her current query and enter a new one. It can be contrasted with search abandonment, which was originally used to describe a situation when a person had given up, or abandoned, their search (Joachims & Radlinski, 2007). Search abandonment was later qualified with the use of ‘bad’ and ‘good’ to distinguish between cases when a person does not click on any search results because none of them are relevant (bad) and when they do not click because they extracted the answer from the result snippet (good) (Li, 2009). For example, a person who enters the query weather and is interested in the current, local weather forecast might not click on any search result if the forecast is displayed on the search results page (good abandonment). Contrast this with a person who enters the same query, but is interested in finding resources to help with writing an essay about the topic of weather. Suppose this person does not click on any result because none of them seem relevant, and then decides to reformulate his/her query (bad abandonment).

Another distinction that can be made between search and query abandonment is that query abandonment allows for instances when a person abandons his/her query even after clicking on one or more search results. While search abandonment takes place after little user engagement with the search engine result page (SERP), with query abandonment searchers can give up on a query after exploring some results first.

Our purpose in introducing these new terms is not to obfuscate things, but to allow for finer-grained distinctions, which we hope will increase theoretical and analytical power. Distinguishing query abandonment from task stopping also allows us to focus on people’s decisions about how much interaction to engage in before they decide to reformulate their queries. It is generally believed that the average web search engine user rarely goes to the second page of search results and there is some empirical evidence to support this (Cutrell & Guan, 2007; Huang, White, & Dumais, 2011; Jansen & Spink, 2005; Lorigo et al., 2008; Spink, Jansen, & Ozmultu, 2000). For example, Spink and colleagues examined queries from an Excite search engine log and noted that people typically examined 1.7 search engine result pages (SERPs) per query and for about 50% of the queries, people went to the next page before reformulating their queries (Spink, Jansen, & Ozmultu, 2000). The researchers also found that the number of instances where only the initial SERP was viewed increased from 29% to 73% in US-based search engines during the period 1997 to 2002 (Jansen & Spink, 2005).

Some recent studies also confirmed that searchers examined fewer than ten results before reformulating by collecting eye movement and mouse hover data. Cutrell and Guan (2007) reported that people examined the first eight results before they re-issued another query. Lorigo et al. (2008) found that participants scanned about 3.2 distinct search results for each query issued. Huang, White, and Dumais (2011) found that people re-queried after inspecting the top four results. Overall, the literature is descriptive in nature; even though some work has attempted to understand the rationales behind good and bad abandonments (Diriye et al., 2012), it has not provided insight into how people make decisions about whether to abandon queries after inspecting the first SERP or paginate to the next SERP page.

Task Stopping

Research about search stopping behavior at the conclusion of a task can be divided into three areas: (1) modeling stopping behavior to inform the development of evaluation measures; (2) understanding how people decide they have found enough information to stop information seeking and (3) identifying cognitive stopping rules that describe when people decide to stop.

Evaluation Measures

Much of the information retrieval (IR) research that has focused on modeling search stopping behavior has done so to inform the development of evaluation measures. As early as 1968, an IR measure called expected search length was proposed as a way to evaluate system performance (Cooper, 1968). Expected search length is the number of non-relevant documents a person has to search through to obtain a desired number of relevant documents; a system with a shorter expected search length is considered as more effective than a system with a longer expected search length. Cooper believed that for any type of search request, specific or exhaustive, there was always a corresponding quantity of relevant documents that the searcher desired. Bates (1984) also argued that the number of results needed depended on whether a task was a high-recall, high-precision or simple search task. Cooper (1973) elaborated two stopping rules for searching ranked output: the frustration point rule, indicating that people stop when a certain number of negative-utility documents are encountered, and the satisfaction stopping rule, by which people stop only when a certain number of positive-utility documents are obtained.

Kraft and Lee (1979) modeled expected search length by using three stopping rules: the satiation rule, the disgust rule, and the combination rule. The satiation and disgust rules were similar to the satisfaction and frustration point rules respectively, while the combination implies that a person stops searching either because of the satiation rule or the disgust rule, whichever comes first. The researchers further demonstrated that expected search length could be approximated using each of the three stopping rules by considering the size of the retrieval set, the number of relevant documents in the set, the number of relevant documents a searcher wished to obtain, and the number of irrelevant documents a searcher would tolerate. In another study related to Cooper’s expected search length, Kantor...
proposed a stopping criterion that incorporated notions of expected probability of success and resilience of a searcher to repeated failures. Kantor proposed that the more a searcher believed he or she would have success in finding a satisfactory document, the more successive non-relevant documents he or she would tolerate. More contemporary IR evaluation measures include parameters to model a searcher’s persistence, which directly reflects stopping behavior [e.g., DCG (Järvelin & Kekäläinen, 2002) and RBP (Moffat & Zobel, 2008)].

**Notions of Enough Information**

Despite the notion that some fixed target number of relevant documents determines when a person stops searching for information, few studies have attempted to determine this number. Many studies instead, have reported that people say they stop searching when they have “enough” information (Zach, 2005). Zach (2005) and Berryman (2006) studied search stopping behavior in non-academic work environments. Zach investigated how senior art administrators determined when to stop searching in their day-to-day jobs and found they mostly stopped either because they felt satisfied with the information obtained or they were forced to stop because of time constraints. Berryman found that public sector policy workers had a difficult time determining how much information would be enough to satisfy their tasks when initiating the tasks, but once the structure was established through information seeking, the stopping point became clearer. This work showed that assessments of enough information are often dynamic and complex.

Some studies have used Herbert Simon’s satisficing (Simon, 1955) to investigate stopping behavior. Satisficing, a decision making process “through which an individual decides when an alternative approach or solution is sufficient to meet the individuals’ desired goals rather than the perfect approach” (Simon, 1971, p. 71) suggests a person stops searching as soon as some certain conditions arise instead of after they have exhaustively considered all available information (March, 1994). Using this theory, Agosto (2002) explored young people’s Web-based decision making. Mansourian and Ford (2007) analyzed the search stopping behavior of staff and students from four Biology departments and classified search stopping behavior by its search depth and search impact. Prabha et al. (2007) investigated the criteria applied by students and faculty when they searched to complete their work tasks. These studies accumulated criteria leading to task stopping in academic information seeking settings.

Only a small number of studies have tried to model or quantify when a person stops searching. Toms and Freund (2009) studied actions preceding end points in online information seeking to predict what actions lead to stopping. The authors found that the most prevalent stopping pattern included issuing a query, examining results and viewing a page. Participants in Dostert and Kelly (2009) stated reasons for stopping searching including intuition that they had enough information, repetition of articles and decrease in relevant articles after conducting assigned study tasks. In an attempt to quantify the feeling of enough information, participants were also asked to indicate how many of the relevant documents they had found. Although participants indicated that they had found 51-60% of the relevant documents, in reality they only found about 7.35%.

**Cognitive Stopping Rules**

Search stopping has also been investigated in the organizational and management science literature, mostly in the context of choice tasks. This research has found that decision makers rely on cognitive stopping rules, or heuristics to make judgments of information sufficiency. Cognitive stopping rules for information search were first introduced by Nickles, Curley and Benson (1995). They identified four rules people use to terminate information search in the context of two house sales prediction tasks and two bank interest rate prediction tasks: (1) mental list: stop when a list of requirements are met; (2) magnitude threshold: stop when the sufficiency of information reaches a predetermined level; (3) difference threshold: stop when not learning anything new; and (4) representational stability: stop when a mental model is formed.

Browne and Pitts (2004) and Pitts and Browne (2004) later investigated the use of these rules by systems analysts during information requirements determination. System analysts were asked to gather requirements until they had enough information to draw diagrams to design an online grocery shopping system. Browne and Pitts found that more experienced analysts tended to use the mental list and magnitude threshold rules, while less experienced analysts applied the difference threshold and representational stability rules. In addition, the application of different stopping rules resulted in varying degrees of quantity, depth, and quality of information. Browne, Pitts and Wetherbe (2007) found that when tasks were complex or when searchers were new to a topic, searchers tended to adopt the magnitude threshold rule or representational stability rule; however, when tasks had low complexity that were easy to decompose into discrete elements, the mental list stopping rule and the single criterion rule play the major roles in triggering search stopping behavior.

**METHODS**

**Data Collection**

A controlled laboratory experimental user study was conducted. The 2-hour study consisted of two stages: (1) online searching and (2) a post-search interview. This paper reports findings from the second stage. Detailed description of the findings from the first stage and the research methods can be found in Wu, Kelly and Sud (2014).
In the first part of the study, participants conducted searches for six open-ended tasks using an experimental search system. All tasks required participants to gather and analyze information, and make a decision. One example task can be found in Figure 1, showing a screenshot of a SERP. The search system was built using the Bing API where the first three query submissions in every task retrieved preselected results to contain different numbers and orderings of relevant and non-relevant documents. Participants’ searches were recorded using Morae.

Figure 1. Example task and Interface

After participants completed all six tasks, a semi-structured, stimulated recall interview was conducted to understand their search strategies for three of the study tasks. Participants were shown the Morae video and were asked questions to understand what they were thinking about as they searched. The purpose of playing the video was to facilitate recall. At the beginning of each play-back for each task, the experimenter started by asking the participants what their first impression was at the sight of seeing the task description. Then the experimenter played the videos until a pagination or reformulation took place on the screen and then asked participants to explain why they paginated/reformulated; this question was intended to collect information about factors contributing to query abandonment. Periodically, participants were also asked why they clicked on certain results. The question about reformulation and pagination was asked repeatedly until the end of each task. Then, participants were asked how they decided it was the time to stop searching. Participants were interviewed using the above procedure for three tasks.

Data Analysis

A combination of deductive and inductive coding was used. Two codes were used initially (deductive): Level and Pattern. These two codes were chosen because level and pattern were independent variables in the first part of the study and we wished to identify evidence that could help explain how they each influenced participants’ query abandonment. Level was associated with SERPs where the number of relevant documents on the first page of results varied, while pattern was associated with the distribution of relevant documents on the first page of results. We also inductively identified codes during the coding process.

Participants

Forty-eight adult participants were recruited from our university by sending an email to the staff mailing list. Because of a logging failure, only interviews with 47 participants are included in this analysis. Seventeen of the participants were male (36%) while 30 were female (64%). Their ages ranged from 19 to 65 and their job titles included web developer, HR specialist, administrative assistant, librarian, lab manager, instructor, fire department technician, and others. Based on the 10-point search self-efficacy scale (Xi & Kelly, 2014), participants held a moderately high level of confidence in their search skills ($M=7.81; SD=1.34$).

RESULTS

The findings are organized as the following: First, we report factors influencing participants’ query abandonment decisions; following that, factors affecting task stopping decisions are discussed.

Query Abandonments

Three factors were found to influence query abandonment: properties of the search results, properties of the queries and properties of the search tasks.

Properties of the Search Results

Participants relied heavily on the first SERP for determining their next moves. The first pages enabled evaluations as to whether a search was “off track” or “on the right track”. In other words, the impression of the first page provided an estimation of the quality of one’s query terms, which was then used to infer the quality of results retrieved by it. Seeing many bad results on the first SERP often caused participants to abandon their queries right away:

They were off track like the tanning one pulled up information about the skin of animals. I knew I was off track so that rather than going to the second page I made more searches. (P4)

One participant used the word “gestalt” to illustrate the overall impression gleaned from the first SERP. Whether to abandon the query was primarily based on the overall appropriateness of the first page of results with respect to what one was searching for:

I guess my rationale was if the things on the first page are not looking so relevant, I probably need a better search term. I am probably not hitting the nail on the head in terms of trying to find what I am finding...It is the overall, gestalt sense, does this seem to fit or not (P43)

While some participants made their query abandonment decisions based on their general and abstract perceptions of the first SERP, others articulated a belief that the more relevant results on the first SERP, the more likely they were to find more relevant results in subsequent SERPs:

I only had three [useful results on the first page], and the rest were about air. I thought if I put in a bunch of words relevant to my query, I can manipulate it to give me some different
results…. If the first page gives me a lot, I will give the second page a chance, but none of the pages gave me a lot. (P18)

In addition, almost every participant made an argument that the most related search results were on the first page, so if the first page, the supposedly best page, did not seem good enough, the likelihood of retrieving good results was low, and there was no point going deeper:

The page I was at already does not have much relevant information. So if I keep going, next, next, the chance of finding more information is less. (P31)

The proportion of relevant results in relation to the proportion of non-relevant results on the first SERP appeared to be another way to measure the quality of the first page and whether to abandon a query:

If every other one is relevant, I might go to the next page...Then I think my phrase is turning up as much bad stuff as relevant stuff, so I am willing to look at more results. (P18)

These comments reported above suggested that the amount of useful information on the first SERP — be it measured by number or proportion of relevant documents, or justified by general impression — affected query abandonment.

Moreover, the relative locations of relevant search results to non-relevant results on a page or the distribution of relevant results or non-relevant on the first page factored into the decision process by some as well:

If there are good ones mixed with irrelevant ones, you can like go on to the second page and find something good, but if it is like good ones and all like junky ones, you might as well move on to the next thing…(P27)

The comment also demonstrated that when several non-relevant results followed relevant results, participants expected seeing more non-relevant results if they kept going; therefore, rather than keep searching they preferred to abandon queries on the first SERPs. Such phenomenon was further elaborated by P17:

As I went further down, it starts to get into leather tanning products. Because I was seeing multiple links like that, it made me think it is probably less likely to find relevance beyond that.

When one participant was asked what he or she would do if the relevant results were evenly dispersed on the first SERP, the participant replied:

If there are relevant pages on the 1st, 3rd, 5th, 7th, I would definitely click on it [the second page]...because then why wouldn't I think that the 9th, the 11th and 13th over the next page be relevant, if it is in that order? (P21)

When relevant results were mixed with non-relevant results, the participant’s belief that search engines ranked results by relevance was undermined so he or she no longer judged what results to view based on the ranks of results. This realization probably caused the participant to believe the pattern observed on the first SERP was the norm, and the pattern should continue in future SERPs. As such, since relevant results were found every other rank in the hypothetical first SERP, the participant was willing to delay query abandonment later with an expectation of retrieving more relevant results.

Query abandonment could also take place when non-relevant search results appeared in a block. Collocating non-relevant search results together made individual non-relevant results even more salient. P22 described how such a result alignment stopped him/her from examining more results:

Something about the five articles, they may be different, but they were all social [stock], social [medial stock], social [stock], the fact they are constructed together just made me think my search terms were not on the head. (P22)

These above comments regarding how the distribution of relevant and non-relevant results influenced subsequent query abandonments demonstrate that the ordering of information was critical to when participants decided to move on to the next query.

Even though participants mostly acknowledged that relevant results should be ranked as early as possible and concluded most useful results should be present on the first SERP based on daily experience, when they believed what they were looking for by nature might have relevant results ranked lower or on subsequent SERPs, they were more motivated to search more before they abandoned their current queries, which to some degree supported the assumption that the more a searcher believes he/she can find a useful result, the more non-relevant results he/she is willing to tolerate (Kantor, 1987). Participants seemed to have developed a knowledge of what type of results were more likely to be ranked higher and what others required more results examination based on their interpretation of the search task and what types of results might interest other people. P27 explained why he/she delayed abandonment:

Because the first page wasn't helping me, I was just going to see if there is anything on the second page. I feel someone could be interested in seeing funny pics so a lot of people clicked on it. (P27)

Sighting seemingly commercial search results sometimes also persuaded participants to search deeper. They believed the results were ranked higher because of payment and that they should skip them (P5). However, others viewed the existence of commercial results as a sign of bad query terms: “When you start to see eBay you know you are in trouble” (P3).

While some participants argued that sifting through search results was laborious, others preferred to stick to searching because they felt, “An extra click is a very small investment for finding something very good.” (P3). Several other participants even exhausted all ten SERPs to make sure they did not miss anything important. One participant
maintained that scrolling down the SERP was so easy that he did not even think he was searching for information, rather, he was letting the computer have the full control, and called such search behavior “lazy search” (P44).

Properties of the Queries
Many participants attributed query abandonment to the nature of their own query terms. While most query abandonments took place early on the first SERPs, they occurred later for reasons such as lack of better terms or lack of confidence in the current query terms. Participants continued examining results until they learned new queries to conduct another search. Participant #3 explained his/her insecurity as the following:

Maybe I am insecure about the terms I use and I want to make sure I exhaust these terms before I change them. If I change before I look at several pages, maybe there is something there. If I change too quickly I am missing out. (P3)

In addition, P3’s comment shows that delaying query abandonment not only was a strategy for looking for useful information but also for looking for alternative query terms. On the other hand, a strong confidence in one’s own query terms also prevented participants from abandoning queries. Participants believed the search results were worth exploring when they were confident their own query terms would retrieve useful information:

...because the search terms are pretty good. They are explicit. So I thought I would expect more results on the first page, but I did not see many, so I thought it would change on the second page. (P30)

Some participants who abandoned queries early attributed their behavior to both the “zero penalty” of reformulations modern search engines afford and to their innate desire to pursue better results:

My going back and retrying, or not going to the next couple of pages maybe isn’t necessarily an indication of impatience, but a knowledge that this is not your last chance....I can come back to this search configuration if I want to....it is a desire, a sense that with a new term maybe I can get better results. (P47)

Another participant argued that abandoning the current query for a new query allowed for new directions of thinking and serendipitous discoveries.

I think it is probably better to submit a new search [when not finding enough on the first page], partially because you are deliberate in your terms that you are searching, you are also putting into a different thought process. It is almost like it allows you to be flexible in your thinking. Because if you are looking for something and you know what the answer is, you are only to get certain amounts of results. And then if you take inspiration by words you see whether it is in an article or a phrase that might lead to a different path, and a deeper understanding about something that maybe related in different ways. (P9)

Properties of the Search Tasks
Participants judged how much effort they should put into searching by task properties. The amount of effort one was willing to put in subsequently influenced how many results they were willing to explore before they abandoned a query. For example, participants’ own task taxonomy allowed them to gauge how much information was needed. Two taxonomies reported by participants were specific vs. opinion and work vs. personal.

If I am looking for something really specific, then I don’t need to, like the fair trade example... I think on things like opinion-based, like the social media question, I always will go to the second page. I mean, if I am looking to see both sides of an argument, something that is vague, something that is opinion-based, then I want to be able to see different things. (P28)

At work I cover my basis more deeply. I need to present information to other people and I don’t want to be caught off guard. I don’t want to be asked questions and oh I wished I had gone deeper. But I feel like at home or things for personal use, I sort of just skim it. (P2)

P2’s and P28’s comments support the view that the amount of information needed to reach the feeling of enough depends on task type (Bates, 1984). P2’s comment also demonstrates that when searching on behalf of other people, there was ambiguity in what enough meant for these people, so he or she adopted higher standards. Yet when searching for oneself, because there were fewer consequential effects, less effort was deemed necessary.

For some participants, they decided to reformulate each time a subtask was fulfilled. Each reformulation marked the division between one search goal and another. As described by Participant #15, “part of the changing of search terms was also to hit the different aspects in the prompt.”

While what P15 said demonstrates that participants read the task descriptions carefully and tried to complete the requirements accordingly, it nevertheless suggests that rather than treating a simulated scenario as a larger task and trying to come up with a search plan on their own, some participants were actually solving multiple independent smaller tasks in a linear order where each subtask appeared in the description. Further, task relevance or task importance determined how much search took place before participants terminated subtasks. For example, Participant #44 said that for something trivial searching at home, 1 or 2 documents was enough; but for things affecting society, he tried to make sure he had critical information by searching for more. When tasks were not important to participants, they did not see the need to search deeply to fulfill them.

Summary
While many participants chose to abandon queries on the first SERP because they believed that search engines present the most useful results on the first SERP or because of the ease of reformulations with search engines, many other abandonments appeared to be contingent on the tasks. The subjective task taxonomies and task relevance allowed
participants to make high-level decisions about the amount of information needed, and the structure of the task influenced the number of query abandonments observed for a search task. The knowledge of the topic of a task gave rise to varying levels of confidence, yet either high or low confidence led to delayed query abandonment for different reasons.

After issuing query terms, more contextual factors influenced query abandonment. The perceived quality of search results, which could be a result of simply the first impression or based on a more careful analysis of the results, had a critical impact on when query abandonment took place. For participants who articulated how specific elements of search results influenced their query abandonment decisions, a lower number of relevant search results was often the reason why they issued another query after examining the first SERP. Even when there appeared to be some good relevant results, when non-relevant results appeared in a block or seemed to occupy the lower end of the first SERP, participants felt continuing to the second SERP was unlikely to be promising. Moreover, the sight of commercial search results played a critical role in participants’ decision-making. Participants also calibrated their search result evaluation efforts based on their assumptions of how search engines responded to third parties. For instance, the knowledge that search engines sold highly-ranked search result space and search results could rank high because many users clicked on them sometimes caused participants to explore search results ranked lower or beyond the first SERP.

**Task Stopping**

Participants tended to explain their task stopping decisions by relating to the content they had reviewed, the goal they were to achieve, how they felt, and the study constraints.

**Content**

A common reason participants stated they stopped searching was because they believed they had exhausted all resources, including search results and known query terms. Participants would not have know objectively whether they had exhausted all relevant search results or not, so the perception of having exhausted results was based on their interactions with search results already seen. One participant explained that he/she stopped because “I could not find anything else and I could not find other terms (P12),” another stopped because the links seemed not relevant anymore (P31).”

Stopping rules identified in previous research were also articulated by many participants. For example, the above phenomenon can be explained with the *Difference Threshold* rule (Nickles et al., 1995). Searchers stop when they do not learn anything new.

Participants did not always push themselves to the extreme; rather, they took the satisfying approach. The concepts of “sufficiency” and “feeling of enough” which have been frequently observed in the literature was repeatedly mentioned during the course of the interview. Participant #43 stated the feeling of enough was: “Much of a qualitative sense, I feel I kind of know what is going on, rather than I know exactly, it's like a gut thing.” Another participant made sure he or she had enough information by asking: “Have I to my satisfaction answered the question? Have I felt I reached a critical mass of the knowledge of this thing?” (P47)

Others took a minimalist approach, believing enough meant “necessary information to answer the question” or “more than one thing, had a variety, and be able to address the question.” Some other participants were able to articulate what enough means in more concrete terms under different circumstances:

> I think it is an important issue; that is why I got four instead of three articles...It is like you have a three or five panel judge. With two [judges] you cannot decide, you got to decide with three, five or seven. (P44)

P44 first explained that for some tasks, three articles normally met the threshold of enough, because three represented the fewest number of voices to reach a diplomatic consensus. Yet in this particular task, which was important to him, four articles rather than three were considered enough.

Not only was the number of articles desired used as a criterion for determining task stopping, for some participants the articles had to belong to a certain genre:

> I guess the source of the documents was from journals, medical standpoint and psychology journals, experts, I thought there was evidence, a good answer from not only someone's opinion. Someone has really done a study... I tried to [monitor how many useful documents I had found]; I was trying to get 4 to 6 documents. (P37)

But for P7, under the circumstance when the documents found were all opinions rather than solid evidence, he or she considered four documents good enough.

> I guess in my searching if I have to read 4 articles, like longer articles about this, then that would be an exhaustive search for me. Because these are people's opinions, I can imagine the 5th 6th 7th articles about the same articles I would sort of get... After the 4th [article] I would probably have formed an opinion...Again, I guess after the 4th one I started to see enough overlap to make me think the fifth one wasn't gonna get me more. (P7)

On the surface it may seem that all these examples supported Kraft and Lee’s (1979) satiation rule and Cooper’s (1973) satisfaction stopping rule since participants stopped after obtaining a certain number of useful documents. Yet the mentioning of how different genres carried different weights demonstrates that quantity alone did not tell the whole story.

Earlier it was mentioned that some participants treated each task as a smaller number of independent tasks. Participants
reformulated once they had fulfilled every subtask listed in the task and they stopped the task once all subtasks were completed. Such a stopping decision can be explained by another rule proposed by the Mental List rule, or stop when a list of requirements are met (Nickles et al., 1995).

In the list of questions I am looking for, I knew I had found several different types of pollutants and I had links saved about how they were harmful, and I felt I had enough relevant ones that I can answer these questions. (P18)

In our study, the requirements were the questions in the task descriptions. The Mental List rule was observed primarily in tasks that appeared to be relatively easy for participants to divide into smaller units, which also supported Browne et al.’s (2007) findings.

In contrast to trying to satisfy each individual subtask, other participants took the approach of obtaining a general understanding of the topic. The Representation Stability rule (Nickles et al., 1995) can be used to explain this approach. Once a mental model was established about the task topic, participants stopped searching for information. An example of stopping following this rule was made clear by Participant #27: “Once I had a good idea, either what I already thought was confirmed, or something new came up, [I stopped].”

At times, participants reported that when a pre-conceived belief was confirmed or supported by retrieved search results, they stopped. Such phenomenon can be explained by the Single Criterion Rule (Nickles et al., 1995). Participants stopped looking once enough information about a single predetermined criterion is found. The single predetermined criterion was often the existing answers or beliefs participants held in mind before searching.

I think [I stopped] because I already had an answer in my mind. I was looking for information to reaffirm mine. I felt like I found several articles that said that. (P24)

Participants who did not have preconceived answers sometimes stopped when they sensed a coherent theme emerging from repetitive information. Participant #43 used the tattoo removal task as an example: “It seems the information was consistent. They all agreed laser was the best. After a couple of research saying that I was like I am not going to find out too much more.”

In tasks where participants were asked to collect information about the pros and cons of using social media and the impact of violent video games, participants were concerned about gathering balanced opinions in order to achieve an unbiased decision. Participant #11 stopped when he/she felt he/she had a good variety of information to look at and it wasn't just one-sided. For Participant #22, a fair assessment of the task required finding a positive, neutral and negative argument.

These two comments, along with the three-panel judge comment mentioned earlier, demonstrate that participants treated debatable topics with greater caution than average open-ended topics where there may exist standardized answers (e.g. most toxic marine pollutant).

It was also found that sometimes participants stopped as soon as a satisfactory document was found. While in many IR models this strategy is hypothesized to occur more often in fact-finding tasks, it seems to apply to open-ended tasks, too. Some participants tended to gather information until they found the best article that met all requirements in the task description. One participant stated that he/she stopped because “this is one of the things there is no answer to it. There is no conclusive evidence, they all say the same things. This one has both sides I feel is better than other sites. (P28)” This happened after participants struggled through repeatedly encountering documents that only fulfilled some aspects of the task requirements but not all.

Goal
Some participants stopped searching once they reached the goals they set out to achieve. One goal participants wanted to achieve with their search was being able to engage in a conversation about the subject matter they had been gathering information on.

By the time I stopped I have managed to find enough sources that gave me 5 or 6 options. And in the hypothetical task if my sister is turning 25 and I have some idea of what she is talking about, then at least I have a basis to have a conversation with her. (P18)

One participant justified the amount of information needed to obtain based on the length of the conversation he or she wished to sustain as the following: “If you and I are meeting for dinner and you brought up this I can easily talk about it for at least 20 minutes (P5).” As an English literature instructor, P5 also imagined he or she was assigning students to write a report about the topic and predetermined a fixed number of resources to be gathered before initiating the search.

Imagine I am writing a paper. Let’s say I make students write a 3 or 4 pages paper investigating in some reasonable manners, 5 to 6 sources, you know, max, so that I will be able to discuss it intelligently at a dinner party, the criteria is to have some reasonable knowledge on both sides and be able to write a 3 to 4 pages paper about it, just to present the facts. (P5)

The strategy P5 took, stopping once a certain amount of information is obtained, fit in the Magnitude Threshold rule proposed by Nickles et al. (1995). Yet it was not commonly observed that participants had a specific number in mind before they embarked on searching.

How They Felt
Frustration from not finding useful information, perceived passage of time, and lack of interest in the topic were some subjective factors that resulted in task stoppings. Participant #6 appeared embarrassed when he or she said: “I was kind of unsatisfied. Like I said, it was really hard to find
conclusive evidence. I figured, well, maybe the fact there is a lot about lasers much less about others tells me maybe laser is it.” Another participant articulated his/her indifference toward the topic:

In a way I just don't care that much about it. I feel like there is so much hearsay about it, like social media is so horrible and ok well I read some of that; that says yes it is bad. And I think after this article I found a couple others that actually have more factual kind of evidence than just opinions. I think that's all I need to know. (P8)

Study Constraints
In addition to lack of interest, which was probably a consequence of administering artificial tasks, participants described several other study constraints that prevented them from spending the time they would have normally spent on searching. Participant #47 put more effort into searching because he/she knew he/she was examined during the study:

None of these are emergency, this is an artificial situation, when I am at home I would bookmark. I can go back. In the spirit of the experiment, I wanted to make sure I tried enough angles. At least have some representative corpus by the time I finished it looked like someone with a brain doing research. (P47)

Participant #15 explained the difference in time between a real need and a simulated need: “If I was truly picking a water sport for my sister's birthday, I would've spent so much more time working. So a lot of it is the time constraint of the study.” Even though no time constraint was placed on participants, P15 was tracking the passage of time to complete all six tasks in a reasonable time. In addition, what could have been a multi-episode task was forced into a single-episode task, which might have also led to the decrease in time and effort participants invested.

SUMMARY
A task stopping decision was often made directly as a result of interactions with SERPs. Participants articulated “forced” stopping, stopping when no more information could be found, or “voluntary” stopping that resulted from securing enough, or necessary information. It was also the case that some participants decided how many results to obtain in the beginning of a task and stopped immediately after the number was satisfied, although this was not a common case. Task structure was used by participants to justify when optimal task stopping should occur just as it was used to justify query abandonment. Moreover, the nature of the search task appeared to influence task stopping decisions as well. For opinion-based tasks, participants focused on gathering a “balance of arguments” and used this to determine stopping. Imagining the potential use of the information one was searching for was also used by some participants to decide when to stop gathering information. Participants related to scenarios where one needed to give others suggestions or one needed to produce a report where concrete criteria needed to be satisfied before quitting. Participants also allowed emotions and motivations to influence their decisions about when to stop. Lastly, participants’ awareness of being in a study motivated some to search more and others to search less.

CONCLUSION AND IMPLICATIONS
Our work investigated two types of search stopping behavior, query abandonment and task stopping, using assigned search tasks in a controlled laboratory environment. Our study is one of the first to probe participants about their decisions to abandon queries, which enhances the numerous studies that report statistics describing the frequencies of such behavior. Our results reveal a variety of factors that affect query abandonment, many of which lead to query reformulations on the first SERP. For participants who had developed the habit of searching only among the first ten results, it was as if only ten, rather than millions of results were retrieved for each query submission. The knowledge of search engine algorithms and the ease of re-querying offered by modern search engines probably also explains why previous studies have found that reformulation is more common than pagination. These findings call for a reconsideration of the current search result presentation practice. What does it mean to offer results that almost no one will examine? How can search results that are not displayed on the first SERP be exposed or integrated to allow for more diverse solutions and serendipity? How can we get people to go deeper in the search results list? And, finally, is there a way we can leverage the habit of early query abandonment to the searchers’ benefits?

With regard to task stopping, our work confirmed the stopping rules proposed by previous work and discovered additional rules. These different stopping rules offered various qualitative and quantitative interpretations of “the feelings of enough”. Our results also aligned with previous findings that different task types required different stopping rules (Browne et al., 2007) and different amount of information to reach the feeling of enough (Bates, 1984). In particular, opinion tasks demanded additional effort in capturing and balancing the “valence” of arguments, or the “pros” and “cons” in the content. This suggests the possibility of displaying search results by the arguments’ stances for opinion-based topics to support learning and decision-making. Offering searchers opportunities to sort retrieved results can also alleviate the burden of constant comparisons between results of different viewpoints.

The use of assigned tasks allowed us to understand the causal relationships between task characteristics and the task stopping rules applied by participants by keeping as many things as constant as possible. However, it changed the time and effort expended by participants and impacted their stopping behaviors. While this can be viewed as a limitation of our study, it is the case that people commonly conduct imposed search tasks, so our findings do have some external validity and suggest that people impose
different stopping rules depending on the origins of the search task.

REFERENCES


