Information Processing and Management 54 (2018) 433–450

Contents lists available at ScienceDirect

Information Processing and Management

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

Examining the impact of domain and cognitive complexity on query formulation and reformulation

Barbara M. Wildemuth[a], Diane Kelly[b,1,*], Emma Boettcher[c,1], Erin Moore[d,1], Gergana Dimitrova[e,1]

[a] University of North Carolina at Chapel Hill, USA
[b] University of Tennessee, Knoxville, USA
[c] University of Chicago Library, USA
[d] Warpwire, Inc., USA
[e] Concordis, USA

A B S T R A C T

The purpose of this analysis was to evaluate an existing set of search tasks in terms of their effectiveness as part of a "shared infrastructure" for conducting interactive IR research. Twenty search tasks that varied in their cognitive complexity and domain were assigned to 47 study participants; the 3,101 moves used to complete those tasks were then analyzed in terms of frequency of each type of move and the sequential patterns they formed. The cognitive complexity of the tasks influenced the number of moves used to complete the tasks, with the most complex (i.e., Create) tasks requiring more moves than tasks at other levels of complexity. Across the four domains, the Commerce tasks elicited more search moves per search. When sequences of moves were analyzed, seven patterns were identified; some of these patterns were associated with particular task characteristics. The findings suggest that search tasks can be designed to elicit particular types of search behaviors and, thus, allow researchers to focus attention on particular aspects of IR interactions.

1. Introduction

Some interactive IR research questions are most appropriately studied in naturalistic settings (i.e., in the wild), but many questions are best investigated through experimental or quasi-experimental studies (Kelly, 2009). In such studies, the researcher will typically assign specific search tasks to be completed by all or a subset of study participants (Kelly & Sugimoto, 2013). Such an approach allows the researcher to control task characteristics across study participants, focusing attention on the independent variables being manipulated.

To be effective in achieving the goals of an experiment, assigned tasks must be designed in a way that is consistent with those goals. For studies that stress ecological validity, simulated work task situations are typically used as the assigned search tasks. "A simulated work task situation is a short textual description that presents a realistic information requiring situation that motivates the test participant to search the IR system" (Borlund, 2016, p.395). By incorporating information about the source of an information need, the environment in which it arose, the problem to be solved, and the objectives of the search, the simulated work task situation can provide study participants with a standardized stimulus for their search and it can thus be assumed that all study participants are

* Corresponding author.
E-mail address: dianek@utk.edu (D. Kelly).
[1] Work conducted while at the University of North Carolina at Chapel Hill.

https://doi.org/10.1016/j.ipm.2018.01.009
Received 14 July 2017; Received in revised form 12 January 2018; Accepted 29 January 2018
Available online 15 February 2018
0306-4573/

responding to the same stimulus. In addition, it can be assumed that the stimulus represents a realistic (even if fictional) work task.

Depending on the researcher's goals, assigned tasks may be designed to elicit particular types of search behaviors (including but not limited to "typical" behaviors). For this reason, tasks may vary by such attributes as type (e.g., Li, 2009; Pharo & Krahn, 2011), complexity (e.g., Albertson & Meadows, 2011; Jansen, Booth, & Smith, 2009), or domain (e.g., Toms, Freund, Kopak, & Bartlett, 2003). In each case, it is assumed that differences in the level or category of each task attribute will have an impact on the search behaviors elicited from the study participants. There is significant evidence to indicate that this assumption is often correct. For example, Jansen et al. (2009) found an impact of the cognitive complexity of the assigned task on the number of queries per session, average query length, unique terms used in the query, as well as other search behaviors. Similarly, evidence for the impact of domain on search behaviors is reported, by Toms et al., (2003). When they compared searches conducted in four domains (consumer health, general research, shopping, and travel), they found effects on the time taken at various search states, the number of instances of each search state, and the search strategies used.

For the study reported here, the cognitive complexity and the domain of the simulated work tasks assigned to the study participants were manipulated. An initial analysis indicated that the task's level of cognitive complexity did have an impact on such searcher behaviors as number of queries generated and their length, number of unique query terms used, number of clicks on the search engine results page (SERP), and number of URLs visited (Kelly, Arguello, Edwards, & Wu, 2015). The additional analyses reported here focus on the effects of the cognitive complexity and domain of the task on the specific search moves made during search strategy formulation and reformulation. This detailed examination of people's search behaviors can potentially reveal how people respond to tasks in particular domains at particular levels of cognitive complexity, and as a consequence may help us understand variations in search behaviors, as well as reveal ways in which assigned search tasks can be designed to elicit particular search behaviors.

2. Background

Three areas of prior research provide a foundation for the current paper. The first focuses on studies of search strategy formulation and reformulation. Past studies have used a number of different approaches to define and analyze search tactics and moves. Here, we will review these approaches, with special attention to their use for examining the sequences of specific moves that constitute a search strategy. Second, we will examine the literature on task complexity. The ways in which complexity has been defined within the context of interactive IR research have recently been reviewed (Wildemuth, Freund, & Toms, 2014), so we will focus attention on studies that have investigated tasks' cognitive complexity. Third, we will review the ways in which the domain of the assigned task may influence search behaviors, particularly when domain knowledge is not expected to vary across study participants.

2.1. Search strategy formulation and reformulation

Wacholder's (2011) review of research on query formulation and reformulation begins by noting the difficulty of studying query formulation processes, since they are cognitive processes and not directly observable. This challenge means that studies of search strategy formulation and reformulation must rely on either the product of the formulation process (i.e., the query that is submitted to the IR system) or the searcher's self-reports of their own cognition. The studies reviewed here, like the current study, based their analyses on the queries directed to the IR system.

To examine the processes of search strategy formulation and reformulation, the researcher must categorize the transitions made from one query to another and/or the multiple search moves represented within each query. Early work taking this approach differentiated search strategies and search tactics, defining a strategy as "a plan for the whole search" and a tactic as "a move made to further the search" (Bates, 1979, p.207). In the following decades, as transaction logs from online search systems became available for analysis, more fine-grained views of searches were possible and researchers began to look at the specific moves made within each query. In her seminal study, Fidel (1985) defined a search move as any "change in a query formulation" made during the course of a search (p. 61). She examined the searches conducted by seven professional searchers and identified eight operational and twelve conceptual moves that they made. Wildemuth & Moore (1995) built on this early work, examining the searches conducted by medical students as they participated in patient care teams.

Others have taken slightly different approaches to defining search moves. Based on a study of searches conducted by veterinary medicine researchers, Shiri & Revie (2003) defined seven cognitive moves and ten physical moves. While this distinction seems parallel to that used by Fidel, Shiri and Revie's cognitive moves are not nearly as fine-grained. A few years later, Rieh & Xie (2006) looked at searchers' reformulations as recorded in Web transaction logs, and categorized them as specification, generalization, replacement with synonyms, or parallel movement (similar to the categories defined by Lau & Horvitz, 1999). After identifying each of the moves made in a search, they focused their analysis on sequential patterns of moves, describing them as search tactics.

Some researchers have augmented their analysis of search moves by including data on the searcher's intentions for a search. Thatcher (2006) collected retrospective think-aloud protocols to more fully understand the Web queries (moves and tactics) generated in response to four assigned search tasks. In a similar vein, (Xie & Joo's 2010; 2012) lab-based study analyzed both the individual moves recorded in the transaction logs and a variety of other interaction-related behaviors (e.g., evaluating retrieved items or monitoring the search process). In both these studies, the use of think-aloud protocols allowed the researchers to infer the searcher's intentions and include those intentions in the definition of particular search moves. This approach, then, will not generalize to studies in which think-aloud protocols are not collected.

Shute & Smith (1993) initiated another stream of research by conceptualizing a search tactic as a "frame-based semantic representation" of the search topic (p.32). Each frame incorporated slots that could be filled with a specific term representing each slot; query reformulations could be changes in which slots were included in the query or changes in the slot fillers. Based on 24 searches conducted by two professional searchers, they identified five "knowledge-based search tactics" (p. 29) for broadening the topic's scope, five for narrowing it, and three for making other types of changes in the topic. While not formally based on Shute and Smith's ideas, the analysis of 1040 Web queries conducted by Bruza & Dennis (1997) is consistent with them. Similarly, Kinley, Tjondronegoro, Partridge, & Edwards, 2014 analyzed 150 Web searches using move categories such as Add, Remove, and Replace. Neither study attempted the more knowledge-based approach of Shute and Smith, since they did not differentiate slots (i.e., concepts) from fillers of those slots (i.e., terms representing the concepts). Wildemuth (2004) adapted Shute and Smith's categories and applied them to the 1298 searches conducted by medical students in response to simulated clinical scenarios. In addition to tabulating the frequency with which particular moves were used, she conducted an analysis of the sequences of moves used together. For the current study, we have adapted the Shute & Smith (1993) knowledge-based search tactics for application in the context of Web searches about everyday information. Our focus is on the ways that people formulate their initial queries and then reformulate those queries during the course of the search.

2.2. Task complexity

Definitions of task complexity often stem from Campbell's (1988) definition of objective task complexity as an attribute of a task having four dimensions: (1) number of paths leading to a desired end state; (2) number of desired outcomes to be achieved; (3) conflicting interdependent relationships between paths to these multiple outcomes; and (4) uncertain or probabilistic links among paths and outcomes. These dimensions are visible in the ways in which researchers have operationalized task complexity in search tasks assigned in interactive IR experiments. For example, Hughes-Morgan & Wilson (2012) assigned complex search tasks that would require query refinement and reformulation, and Bell and Ruthven (2004) assigned complex search tasks for which both the process of searching and the desired outcomes were unclear. The various ways in which search task complexity has been operationalized are reviewed by Wildemuth, Freund, and Toms (2014).

As outlined by Byström & Hansen (2005), consideration of tasks must be multi-layered. At least within the context of an organization, work tasks are "separable parts of a person's duties to her/his employer" (Byström & Hansen, 2005, p.1053). When considering everyday life situations, "daily-life tasks" are a parallel aspect of the context of information seeking (Ingwersen & Järvelin, 2005, p.32). Here, we will consider the layer of tasks that motivates and guides information seeking and information retrieval tasks and, for convenience, refer to them as work tasks.

Early work focusing on the complexity of work tasks was conducted by Byström & Järvelin (1995). They defined complexity from the perspective of the person completing the task, as the "*a priori* determinability of, or uncertainty about, task outcomes, process, and information requirements" (p.194) and developed a five-part taxonomy based on the single dimension described in their definition of task complexity. Li & Belkin (2008) developed a multi-faceted taxonomy of task complexity, taking into account such "generic" task facets as their source, their frequency, and the type of product that will be the outcome of task accomplishment, among others. This taxonomy provides a theoretical framework for further studies of the relationships among work tasks and the search tasks they motivate.

Taking a somewhat different perspective, Jansen et al. (2009) conceptualized searching as embedded within a learning process, rather than within a decision making or problem solving process. This perspective led them to apply Anderson & Krathwohl's (2001) version of Bloom's (1956) taxonomy of cognitive processes to develop simulated work task scenarios at each of the six levels: Remembering, Understanding, Applying, Analyzing, Evaluating, and Creating. For example, a scenario at the Understanding level must require study participants to "translate, construe, interpret, or extrapolate information", while a scenario at the Evaluating level must require participants to "appraise or relate information to the real world" (p.649). A similar approach was taken by Kelly et al., (2015) and also formed the basis for the current study. Based on these prior studies, there is some evidence that work tasks based on a higher level of cognitive complexity also elicit more effort from study participants. Jansen et al. (2009) found a curvilinear relationship, with Applying and Analyzing tasks eliciting more queries, longer queries, more unique search terms, longer search sessions, and more results pages viewed than tasks at either end of the spectrum of cognitive complexity. Kelly et al., (2015) found a positive relationship between the cognitive complexity of the task and user effort for total time spent, number of queries submitted, SERP Clicks, and number and diversity of URLs visited. However, in this study, tasks that were associated with higher levels of cognitive complexity were not rated as more complex or difficult by participants, which suggests that cognitive complexity and task complexity are separate theoretical constructs. Discrepancies between results from Jansen et al. (2009) and Kelly et al., (2015) suggest that more in-depth analysis of the search strategies used may help us understand the influence of a task's cognitive complexity on search behaviors.

2.3. Domain

As noted by Monchaux, Amadieu, Chevalier, & Mariné, (2015), "prior domain knowledge is a major predictor of performance in information searching" (p. 557). However, few studies have examined the influence of domain knowledge on the search processes and strategies that people use; the available studies have been reviewed by Vibert et al., 2009, Wildemuth (2004), and Zhang, Liu, Cole, & Belkin, (2015). While the search tasks in the current study vary across four different domains, its focus is not on the effects of differing levels of domain knowledge on search moves; instead, it is on the possible effects of domain on search moves when domain knowledge is not expected to vary.

Many studies incorporate a set of assigned search tasks that originate in multiple domains. For example, Liu, Kim, & Creel, (2015) assigned tasks about soccer and philosophy, among others, and Kinley et al., (2014) assigned tasks on child safety laws in Texas, trekking in Nepal, and the Bermuda Triangle mystery. Liu et al., (2010) provided a single context – the profession of journalism – for simulated work task scenarios dealing with different topics, including US visa laws, South Korean presidential politics, state budget cuts in New Jersey, and a particular artist. Unless differences in participant domain knowledge are being investigated, the purpose of diversity in task topics/domains is to elicit a more representative sample of search behaviors and the effects of the search task domain are implicitly assumed to be negligible.

Some studies (e.g., Toms et al., 2003) have directly examined the impact of the domain on people's search behaviors, including the efficiency and effectiveness of the search, and the specific search strategies used. In other studies, a single domain was investigated and the focus was on the impact of the searcher's level of domain expertise. For example, Karimi, Scholer, Clark, & Kharazmi, (2011) investigated the effects of expertise in the biomedical domain on such search behaviors as the number of queries entered, the types of queries used, the results pages viewed, and the items saved. Such studies, where domain is a key aspect of the study design, demonstrate empirically that the domain of the assigned search task cannot be ignored.

Although neither Kim & Soergel's (2005) or Li & Belkin's (2008) faceted classifications of task characteristics take task topic or domain into account, others have argued that task domain is an important consideration. Kelly (2009) points out that an important dimension of a person's information need is the topic, which "represents the subject area that is the focus of the task" (p.79). Taylor's (1991) explication of information use environments notes that the domain of interest is one aspect of the setting in which an information need can arise. Taylor argues that each domain "will have certain attributes peculiar to that domain: availability of information, patterns of dissemination, and to some extent the level of reliability" of the information available (p. 227). In addition, the problems that arise may vary by domain. While Taylor is not solely concerned with the "subject matter as a definer of problems" (p.224), he does acknowledge its role in defining something as a problem and specifying appropriate resolutions of those problems. Supported by Taylor's framework, our argument is that particular domains generate and recognize particular types of problems, many of which are resolved through completion of work tasks. For interactive IR studies, researchers need to take into account the role of domain in shaping the search tasks they assign. Thus, the current study investigates the impact of domain on search moves.

2.4. The research problem

The focus of the current study is on search strategy formulation and reformulation, as represented in specific search moves. It is a secondary analysis of data collected in a previous study (Kelly et al., 2015). The goal of this previous study was to develop and evaluate a set of search tasks that would elicit a range of search behaviors, which could be used in interactive information retrieval evaluations. Kelly et al., (2015) created a set of 20 search tasks using a framework that consisted of four domains and five levels of cognitive complexity. In their analysis, they focused on search behaviors at the search task (or session) level; that is, they considered aggregate measures of behaviors such as time taken to complete the task and number of queries issued during the task, but did not analyze the specific moves people made while completing the tasks. In this paper, we present a fine-grained analysis of the sequences of individual search moves incorporated in searches. The effects of the cognitive complexity of tasks on search moves are examined, as well as the effects of domain. In addition, the implications of these effects for the design of search tasks will be discussed in order to provide guidance for interactive IR researchers in planning future studies.

3. Methods

Forty-eight research participants each completed five assigned search tasks, one from each of five different levels of complexity. Across participants, the tasks were derived from four different domains; each participant searched in only one domain. The individual search moves were coded using a coding scheme adapted from Shute & Smith (1993). The search moves were analyzed in terms of their overall frequency of occurrence, and differences in frequency across domains and complexity levels. In addition, two types of sequential analyses of the search moves were performed. The study methods are described in more detail below.

3.1. Sample

Forty-eight undergraduates participated, including 33 females and 15 males, of an average age of 20 years (SD = 1.62). Their majors included the sciences (n = 10), the social sciences (n = 28), the humanities (n = 3), a profession (n = 6), and one undecided. Most reported conducting searches daily and had been searching for 7–9 years. Transaction logs of the searches were captured. Because of a logging failure, data from only 47 participants are included in the analyses reported in this paper.

3.2. Search tasks

Sets of tasks were created in four domains believed to be of general interest to the target audience: Commerce, Entertainment, Health, and Science & Technology. Each task set included one task at each of five complexity levels, based on Anderson & Krathwohl's (2001) taxonomy of educational objectives. The levels used in the study were: Remember (retrieve, recognize, and recall relevant knowledge), Understand (construct meaning through interpreting, exemplifying, classifying, summarizing, inferring, comparing, and explaining), Analyze (break material into constituent parts, determining how the parts relate to one another and to an overall structure or purpose through differentiating, organizing, and attributing), Evaluate (make judgments based on criteria and standards

Table 1
Example search tasks.

Complexity level	Domain	Search task
Remember (level 1)	Health	You recently watched a documentary about people living with HIV in the United States. You thought the disease was nearly eradicated, and are now curious to know more about the prevalence of the disease. Specifically, how many people in the US are currently living with HIV? *Concepts/facets: HIV, living, United States, population/statistics, current/currency*
Understand (level 2)	Commerce	You have noticed that some coffee shops in your neighborhood advertise that they only sell 'fair trade' coffee. In order to decide whether to support these coffee shops you want to understand what the label 'fair trade' really entails. What are the requirements for coffee to be labeled as fair trade? *Concepts/facets: fair trade, coffee, requirements for labelling*
Analyze (level 3)	Entertainment	Your sister is turning 25 next month and wants to do something exciting for her birthday. She is considering some type of extreme sport. What are some different types of extreme sports in which amateurs can participate? What are the risks involved with each sport? *Concepts/facets: extreme sports (including specific sports), amateur, risks, types/lists*
Evaluate (level 4)	Science & Technology	You have noticed that online services such as Facebook have replaced face-to-face interactions. You can see the advantages of this style of communication, but your sibling argues that people are losing their ability to communicate face-to-face. In general, does use of computers for communication have a positive or negative impact on people's face-to-face social skills? *Concepts/facets: computers/technology, communicative uses of technology, impact/effect (including both positive and negative, face-to-face social skills*
Create (level 5)	Health	Your great granny's doctor has told her that getting more exercise will increase her fitness and help her avoid injuries. Your great granny does not use the Internet and has asked you to create an exercise program for her. She is 90-years old. Put together two thirty-minute low-impact exercise programs that she could alternate between during the week. *Concepts/facets: exercise, low-impact, 30 minute (time), routine/program, alternate/differentiate, elderly/granny*

through checking and critiquing), and Create (put elements together to form a coherent or functional whole; reorganize elements into a new pattern/structure through generating, planning, or producing). An example task at each level is provided in Table 1, along with the concepts, or facets, explicitly included in the task description. The full set of tasks can be viewed online at http://ils.unc.edu/searchtasksforiir/.

3.3. Data collection procedures

Each participant completed five search tasks (one representing each cognitive complexity level) from a single domain. Tasks were rotated using a Latin-square. Participants were presented with each search task, which was printed on a piece of paper, and asked to search the open Web using the search engine of their choice (in all cases, Google). Participants were asked to create responses to each task by typing answers and/or copying and pasting evidence that helped them arrive at their answers into a word processing document. No task time limits were imposed. Before and after each search task, and at the end of the experimental session, participants completed questionnaires about their experiences (not analyzed here).

3.4. Data set

Table 2 displays data describing participants' search interactions, by cognitive complexity level. Time is reported in minutes and query length is the number of words in the query, including stop words. SERP Clicks are the number of pages to which participants navigated from the search engine result page, and Browse moves represent the number of non-SERP pages visited through, for example, following hyperlinks. These data include 235 searches (47 participants × 5 domains). These data are included here to provide context; interested readers are referred to Kelly et al., (2015) for more details, including analyses of differences in interaction measures by cognitive complexity level.

In the current paper, the raw data set was composed of the queries submitted for each search and the URLs visited, either from the

Table 2
Mean (standard deviation) search interactions, by cognitive complexity level.

	Cognitive complexity level*				
	1	2	3	4	5
Time (min)	2.84 (2.45)	5.47 (3.38)	8.34 (4.30)	9.25 (5.27)	9.87 (5.30)
Queries	1.68 (1.05)	1.87 (1.06)	2.94 (1.90)	2.51 (2.25)	4.85 (4.42)
Query length	6.01 (2.94)	5.50 (2.30)	4.68 (2.81)	5.39 (3.21)	4.04 (1.37)
SERP clicks	2.49 (1.56)	2.87 (1.61)	3.64 (2.67)	3.77 (4.16)	5.98 (5.02)
Browse moves	3.70 (3.93)	4.98 (3.99)	6.49 (4.43)	7.70 (8.05)	14.43 (12.34)

*Cognitive complexity levels: 1 = Remember, 2 = Understand, 3 = Analyze, 4 = Evaluate, 5 = Create.

Table 3
Coding scheme.

Code	Definition	Example
Exhaust	All the facets of the task are included in the first query of a task; Note: length varies by the number of facets in the assigned task	Q1: long-term health risks for high school football players
Initial Concept	The first concept observed in the first query submitted for a task	
Add Concept	Add a concept that is not represented in the previous query/move; may include a concept not explicitly included in the search task	Q1: black lamp Q2: black lamp buy
Add Term	Add a term leading to a narrower specification of the concept	Q1: baseball mascot Q2: unique baseball mascot
Narrow Term	Replace a term with a narrower term for the same concept	Q1: artificial tanning Q2: spray tanning
Delete Concept	Broaden the search by deleting one of the concepts included in the previous query	Q1: high school football player Q2: football player
Delete Term	Remove a query term while retaining the original concept	Q1: modern living room rug Q2: rug
Broaden Term	Replace a term with a broader term for the same concept	Q1: gaming laptop Q2: gaming computer
Replace Term	Replace a term with a sibling or cousin term (i.e., a synonym or closely-related term) for the same concept; does not necessarily require a 1:1 equivalence in number of words	Q1: exercise activities for seniors Q2: sports for seniors
Edit	Make minor (i.e., non-conceptual) changes in the query; can include changes in word order, respacing, correction of spelling errors, and adding or removing stopwords	Q1: longterm football injuries Q2: long term football injuries
Error	The entire query is considered an error (and it was skipped in terms of analyzing transitions to the next query)	Q1: mascotts of the international league
SERP No Click	Display of a search engine results page (SERP) with no clicks made on page	
SERP Click	Click on a link to a web page listed on the search engine results page (SERP)	
Browse	Use a hyperlink on a viewed web page to move to another web page	

search results page or through browsing from another page. The transaction logs were coded, using a scheme adapted from Shute & Smith's (1993) knowledge-based coding scheme (see Table 3). Prior to coding, each task was analyzed in terms of the concepts explicitly mentioned in the task description (Table 1). In this way, new concepts could be seen as they entered the query, and differentiated from instances in which the terms representing a single concept were being changed. This approach to the coding also allowed the coders to identify instances in which the study participants introduced concepts that were not explicitly mentioned in the task description. A single query often included more than one move, e.g., the query "bcs football" was coded as an Initial Concept move (for the concept, bcs) and an Add Concept move (for the concept, football). In addition, display moves were coded as SERP No Click, SERP Click, or Browse moves. Three of the authors independently coded the queries; any differences in coding were reconciled through discussion among all five authors.

3.5. Data analysis

Prior to analysis, some adjustments were made to the data set. First, Edit and Error moves were removed. These tended to be typographical errors that were quickly remedied, and occurred infrequently (n = 44). Since they did not represent the cognitive work of responding to the search task, they were eliminated from further analysis. Second, the 40 Exhaust moves were decomposed into their 137 constituent Initial/Add Concept moves. This change was made because the coding of Exhaust moves is very dependent on the number of facets in the search task assigned. However, because the Exhaust moves do represent a particularly interesting type of search tactic, they are analyzed and discussed separately.

After preparation of the data set, 3101 moves (including both query moves and display moves) were identified across the 235 searches conducted by the 47 participants. The moves were analyzed in terms of their frequency of occurrence: overall, by search task domain, and by cognitive complexity level. The results contingency tables were evaluated using chi-square, with alpha set at 0.05. When statistically significant chi-square results were found, the results were further evaluated based on the standardized residuals for each cell to more specifically pinpoint significant differences (Agresti, 2002; Sharpe, 2015).

The sequences of moves in each search were also analyzed. For this sequential analysis, two types of searches were removed from the data set. First, 14 extremely long searches (composed of a total of 597 search moves) were removed. To identify these searches, all searches consisting of 30 or more moves were examined. If a preponderance of those moves were display moves, the search was retained in the analysis, since the queries and their reformulations were comparable to the other searches in the data set. If the search included a high proportion of query moves, it was excluded from the analysis. This operationalization of an extremely long query is consistent with the method used by Hassan, White, Dumais, & Wang, (2014) to filter their data set by query length. These 14 searches represented the use of multiple tactics (each comprised of multiple moves), but there were too few of them to identify patterns in the way tactics were combined. These extremely long searches are presented and discussed in a separate section of the results. Second, 107 single-query searches (composed of a total of 719 search moves) were removed. These searches do not represent query reformulations, since only a single query was entered, and the query moves in this set are composed entirely of Initial/Add Concept moves. Thus, they did not add to our understanding of the ways in which people reformulate their search strategies across queries. In total, the sequential analysis incorporated 114 searches composed of 1698 moves. This smaller data set was analyzed in two ways.

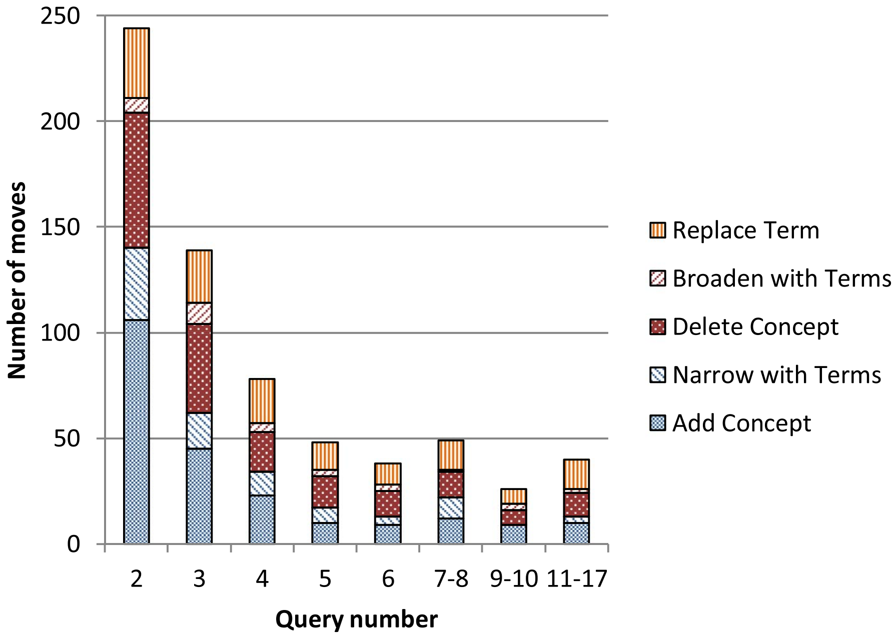

Fig. 1. Types of moves over the course of a search, excluding first query.

First, a first-order state transition matrix was constructed to examine the frequencies of the transitions from one move to the next. Second, a qualitative analysis was conducted, using each full search as the unit of analysis. In this analysis, similar search strategies were grouped into search strategy patterns.

4. Results

4.1. Frequency of moves

A total of 3101 moves were made when completing the search tasks. Thus, there were 13.2 moves per search, on average. Fig. 1 displays the types of moves over the course of the queries submitted for a search. The moves made during Query 1 are not shown, since they included only 235 Initial Concept moves and 385 Add Concept moves; of more interest is the distribution of moves made in subsequent queries. For example, the first and tallest bar shows the distribution of moves that were made by participants during the second query they issued for a task. It includes all the different types of moves (other than Initial Concept) and all types of moves continued to be present through all the subsequent queries.

Table 4 displays the mean number of moves per search, by cognitive complexity level. The mean number of search moves increased as cognitive complexity increased, with participants' searches for Create (level 5) tasks being nearly four times as long as their searches for Remember (level 1) tasks. An ANOVA showed a significant difference in these means, F (4, 234) = 21.44, $p < 0.0001$. Bonferroni's follow-up test results indicated that Create (level 5) tasks elicited significantly longer searches than any of the other types of task.

Table 5 displays the mean number of moves per search, by domain. An ANOVA showed a significant difference in these means, F (3, 46) = 11.93, $p < 0.0001$. Bonferroni's follow-up test results indicated that Commerce tasks elicited significantly longer searches than any of the other domains.

Forty-two percent (n = 1282) of all the moves were query moves and 58% (n = 1819) were display moves. Table 6 shows the frequencies of different moves and the distribution of these moves, by cognitive complexity. Among the seven types of query moves, Add Concept occurred with the greatest frequency, accounting for 20% of the query moves. Initial Concept accounted for 8% of the

Table 4
Mean (standard deviation) number of moves per search, by cognitive complexity.

Cognitive complexity level	Mean	Standard deviation	Minimum	Maximum
Remember (1)	6.94	3.77	2	20
Understand (2)	9.47	5.07	4	32
Analyze (3)	12.68	6.39	2	29
Evaluate (4)	12.57	9.74	4	49
Create (5)	24.32	17.63	3	65

Table 5
Mean (standard deviation) number of moves per search, by domain.

Domain	Mean	Standard deviation	Minimum	Maximum
Commerce	102.55	28.74	60	138
Entertainment	54.83	21.15	27	93
Health	60.67	28.03	27	129
Science & technology	48.92	14.31	29	83

Table 6
Frequency distributions of moves, by cognitive complexity level.

	Cognitive complexity level*					Total
	1	2	3	4	5	
QUERY moves						
Initial Concept	47	47	47	47	47	**235**
Add Concept	83	128	110	128	160	**609**
Add Term	4	6	6	3	24	**43**
Narrow Term	1	1	20	2	19	**43**
Delete Concept	6	19	32	28	97	**182**
Delete Term	1	0	3	4	14	**22**
Broaden Term	1	2	1	5	2	**11**
Replace Term	9	7	40	23	58	**137**
DISPLAY moves						
SERP No Click	9	12	26	19	68	**134**
SERP Click	107	135	184	180	248	**854**
Browse	58	88	127	152	406	**831**
Total	**326**	**445**	**596**	**591**	**1143**	**3101**

*Cognitive complexity levels: 1 = Remember, 2 = Understand, 3 = Analyze, 4 = Evaluate, 5 = Create.

moves; its occurrence was uniform across cognitive complexity levels because this code was consistently used for the first concept observed in the first query for each task. Delete Concept (6% of all query moves) and Replace Term (5% of all query moves) were also used frequently. With respect to display moves, there were roughly equal occurrences of SERP Clicks and Browse moves (46% and 47%, respectively, of the display moves). SERP No Click comprised about 7% of all display moves.

The 40 Exhaust moves were decomposed to the appropriate number of Initial Concept and Add Concept moves prior to this analysis, as described above. An Exhaust move was used for only seven of the 20 search tasks, and almost exclusively for Remember (level 1) and Understand (level 2) tasks. These task descriptions were the most specific and detailed since they asked participants to find a fact (e.g., what is the name of the deepest part of the ocean?) or a basic description of a phenomenon (e.g., how are football BCS rankings determined?). This move occurred only once for an Evaluate (level 4) task and did not occur at all in response to Analyze (level 3) or Create (level 5) tasks. At these higher levels of cognitive complexity, more general terms and concepts were used in the task descriptions and the number of possible solutions was large and mainly dependent on how the participant decided to proceed.

To support a valid chi-square test of the possible relationship between the frequency of the search moves and the task complexity level, some of the less frequently used moves were grouped. Add Term and Narrow Term were grouped, since these two types of term manipulations both had the effect of decreasing the size of the retrieved set. Analogously, Delete Term and Broaden Term were grouped as moves that would increase the size of the retrieved set. Using this grouping, move frequency by cognitive complexity level is shown in Table 7. In general, the number of moves of each type increased with increases in task complexity. More concepts and terms were added, deleted and replaced for more cognitively complex tasks; there were also increases in the frequencies of the two types of display moves as cognitive complexity increased. This pattern held true for all types of moves across all levels of complexity, with the occasional exception of a dip in the number of moves used to complete Analyze (level 3) or Evaluate (level 4) tasks, and of Initial Concept, which was uniform across complexity levels due to the way it was defined. A chi-square test was used to examine distributions of the grouped search moves across cognitive complexity levels. Initial Concept was not included in this analysis because it was not possible for the observations to vary across complexity level. Results of this test showed a significant difference in these distributions, $X^2(28, 2866) = 207.25$, $p < 0.0001$.

The standardized residuals are also shown in Table 7. The magnitude of the standardized residual indicates the extent of the difference between the expected and observed counts; a standardized residual that exceeds about two or three in absolute value indicates a lack of fit of the null hypothesis in that cell (Agresti, 2002, p. 81). This analysis can be used to understand the move profile for tasks at different levels of complexity by examining the columns of Table 7. For Remember (level 1) tasks, a greater proportion of moves were Add Concept and SERP Click, while Delete Concept and Browse were observed less often than expected. For Understand (level 2) tasks, Add Concept was observed more often than expected, while Replace Term and Browse were observed less often. For Analyze (level 3) tasks, Narrow Search with Term Manipulation and Replace Term were observed more often than expected, while Browse was observed less often than expected. For Evaluate (level 4) tasks, the proportions for all moves were as expected except for

Table 7
Counts and standardized residuals of moves, by cognitive complexity level.

	Cognitive complexity level*					Total (%)	
	1	2	3	4	5		
Initial Concept	47	47	47	47	47	**235**	***7.6***
	–	–	–	–	–		
Add Concept	83	128	110	128	160	**609**	***19.6***
	3.1	*4.7*	*−0.6*	*1.2*	*−4.8*		
Narrow search with term manipulation	5	7	26	5	43	**86**	***2.8***
	−1.5	*−1.7*	*2.4*	*−2.7*	*2.1*		
Delete Concept	6	19	32	28	97	**182**	***5.9***
	−2.8	*−1.2*	*−0.5*	*−1.1*	*3.3*		
Broaden search with term manipulation	2	2	4	9	16	**33**	***1.1***
	−0.8	*−1.3*	*−0.9*	*1.0*	*1.2*		
Replace Term	9	7	40	23	58	**137**	***4.4***
	−1.2	−2.8	2.7	−0.6	.8		
SERP No Click	9	12	26	19	68	**134**	***4.3***
	−1.1	*−1.5*	*0.1*	*−1.3*	*2.3*		
SERP Click	107	135	184	180	248	**854**	***27.5***
	2.6	*1.5*	*1.6*	*1.4*	*−4.3*		
Browse	58	88	127	152	406	**831**	***26.8***
	−2.5	*−2.6*	*−2.6*	*−0.5*	*4.9*		
Total	**326**	**445**	**596**	**591**	**1143**	**3101**	***100.0***

*Cognitive complexity levels: 1 = Remember, 2 = Understand, 3 = Analyze, 4 = Evaluate, 5 = Create.

Narrow Search with Term Manipulation, which was observed less often than expected. Finally, for Create (level 5) tasks, Narrow Search with Term Manipulation, Delete Concept, SERP No Click and Browse were observed more often than expected, while Add Concept and SERP Click were observed less than expected.

Table 8 shows the distribution of moves by task domain. Overall, participants executed the most moves when completing commerce tasks; the number of moves used in searches of the other three task domains were roughly equal. A chi-square test showed significant differences in the distribution of moves between domains, $X^2(21, 2866) = 216.53$, $p < 0.0001$.

The results were further evaluated based on the standardized residuals. Examining the column data allows us to understand each domain's move profile. For Commerce tasks, participants made fewer Add Concept moves and SERP Clicks than expected, and more Delete Concept, Replace Term and Browse moves than expected. For Entertainment tasks, participants made more Narrow Search with Term Manipulation moves and fewer Browse moves than expected. For Health tasks, participants made fewer Delete Concept and Replace Term moves than expected. Finally, for Science & Technology tasks, participants made more Add Concept moves and SERP Clicks than expected and fewer Narrow Search with Term Manipulation and Browse moves than expected.

The relationships between cognitive complexity and task domain were also examined in terms of overall move frequency. A chi-square test detected a significant difference in these distributions, $X^2(12, 3101) = 105.71$, $p < 0.0001$. The results were further

Table 8
Counts and standardized residuals of moves, by domain.

	Commerce*	Entertainment	Health	Science & technology	Total (%)	
Initial Concept	55	60	60	60	**235**	***7.6***
	–	–	–	–		
Add Concept	182	134	147	146	**609**	**19.6**
	−3.0	*0.6*	*0.4*	*3.2*		
Narrow search with term manipulation	29	30	21	7	**86**	***2.8***
	−0.5	*2.5*	*0.2*	*−2.2*		
Delete Concept	88	47	24	23	**182**	***5.9***
	2.4	*1.5*	*−2.8*	*−1.8*		
Broaden search with term manipulation	7	9	8	9	**33**	***1.1***
	−1.5	*0.8*	*0.1*	*1.2*		
Replace Term	66	34	20	17	**137**	***4.4***
	2.1	*1.0*	*−2.1*	*−1.6*		
SERP No Click	63	29	23	19	**134**	***4.3***
	1.8	*0.2*	*−1.5*	*−1.1*		
SERP Click	220	194	222	218	**854**	***27.5***
	−5.6	*1.2*	*1.6*	*4.9*		
Browse	419	121	203	88	**831**	***26.8***
	6.1	*−4.0*	*0.7*	*−5.2*		
Total	**1128**	**658**	**728**	**587**	**3101**	***100.0***

*There are five fewer observations for the Commerce tasks because of a logging failure.

Table 9
State transition matrix of moves.

Move	Next move									
	Add Concept	Narrow with terms	Delete Concept	Broaden with terms	Replace Term	SERP No Click	SERP Click	Browse	End	Total
Initial Concept	**104**					2	7	1		**114**
Add Concept	81	13	1	2	11	43	**150**	1		**302**
Narrow with terms	3	2			7	16	35	1		**64**
Delete Concept	40	14	23	2	6	7	9			**101**
Broaden with terms	3	1		2	1	2	12			**21**
Replace Term	14	1		2	8	16	43	1		**85**
SERP No Click	20	11	17	3	18				17	**86**
SERP Click	23	12	32	9	26		**195**	**124**	52	**473**
Browse	14	10	28	1	8		22	**411**	45	**539**
Total	**302**	**64**	**101**	**21**	**85**	**85**	**473**	**539**	**114**	**1785**

Note: Those transitions that make up 5% or more of all the transitions (89+) are in bold and underlined; those that make up 1.5% or more of all transitions (27+) are underlined only.

evaluated based on the standardized residuals. In the Commerce domain, the Evaluate (level 4) task elicited more search moves than expected, while the Remember (level 1), Understand (level 2) and Analyze (level 3) tasks elicited fewer moves than expected. In the Entertainment domain, the Analyze (level 3) task elicited more than the expected number of moves, while the Evaluate (level 4) task elicited less than the expected number of moves. In the Health domain, Remember (level 1) tasks elicited more moves than expected. In the Science & Technology domain, the Understand (level 2) task elicited more than the expected number of search moves, while the Create (level 5) task elicited fewer moves than expected.

4.2. Moves as state transitions

Our analysis of the sequences of moves included in the searches began with a focus on pairs of moves (i.e., transitions from one state/move to the next). For this analysis, 114 of the 235 searches were included; the single-query searches and a few very long searches were excluded, as explained earlier. A first-order state transition matrix is presented in Table 9. The matrix shows the dominance of just a few transitions: Initial Concept to Add Concept, Add Concept to SERP Click, SERP Click to SERP Click, SERP Click to Browse, and Browse to Browse. These findings are consistent with the overall move frequencies shown in Table 6, where the Add Concept, SERP Click and Browse moves each occurred over 600 times in the full data set.

A further examination of the important but less frequent transitions (accounting for 1.5% or more of all the transitions) is also useful. We can see that an Add Concept move is likely to be followed by another Add Concept move or a SERP No Click move. When a Narrow with Terms move is used, it is most likely to be followed by a SERP Click. When a Delete Concept move is used, the most likely next move is an Add Concept move or another Delete Concept move. When a Replace Term move is used, it is most likely to be followed by a SERP Click. Viewing the SERP without clicking is most likely to be followed by an Add Concept move or a Replace Term move. In addition to being followed by other display moves, a SERP Click is likely to be followed by an Add Concept move, a Delete Concept move, or a Replace Term move. Other than another Browse move, a Browse move is likely to be followed by a Delete Concept move or a SERP Click. This analysis provides a better view of the sequential process embodied in these searches than simple move frequencies provide, but it still looks at only two moves at a time. Therefore, an additional qualitative analysis was conducted.

4.3. Search strategies

As noted in the Methods section, this analysis used the full search as the unit of analysis, rather than the individual moves (as in the first analysis) or transitions between pairs of moves (as in the second analysis). It included 114 searches, after the single-query searches and the extremely-long searches were removed. Only seven different patterns of search strategies occurred among these 114 searches, shown as Patterns A-G in Fig. 2.

Pattern A is Narrow–Display–Narrow–Display, and incorporated 25 searches. Essentially, the searcher was working to narrow the retrieval set throughout the search. Searches in this group incorporate 1-2 Add Concept moves after the Initial Concept move (prior to a display). After displaying some results or at least viewing the search engine results page (SERP), the searches incorporate 1-2 more Add Concept moves and/or 1-2 Narrow Term moves. These were followed by a final set of display moves.

Pattern B is Narrow–Display–Broaden–Display. The 13 searches using this pattern begin just like the searches in Pattern A, with 1-4 Add Concept moves after the Initial Concept move. As shown in Fig. 2, there are few Browse moves in Pattern B, and some participants go directly to broadening the search without clicking on any of the search results. Broadening the search is accomplished through a combination of 1-3 Delete Concept moves and/or 1-2 Broaden Term moves. This group of strategies ended with a few more SERP Click and/or Browse moves.

Pattern C is different from Group B in that it adds another cycle of narrowing the search before concluding with the final display

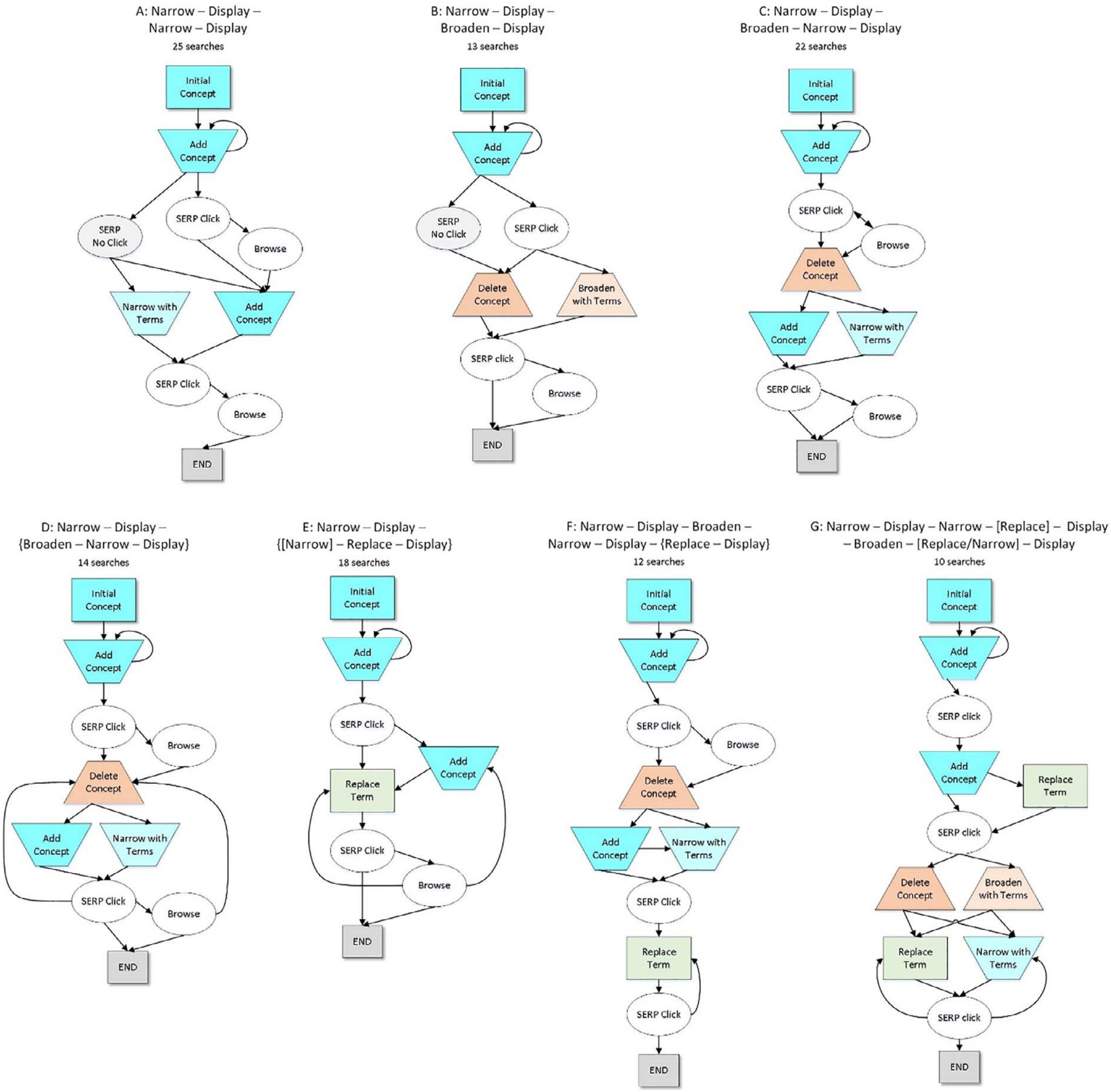

Fig. 2. Search strategies.

moves: Narrow–Display–Broaden–Narrow–Display. In these 22 searches, there are only one or two Add Concept moves after the Initial Concept move. After displaying some results, the strategies using this pattern broaden the search, using 1-2 Delete Concept moves or a Delete Term move. Without clicking through to any of the results and sometimes within the same query, these strategies then narrow the search, using 1-2 Add Concept moves, an Add Term move, and/or a Narrow Term move. The searches using this pattern conclude with a mix of SERP Click and Browse moves.

Pattern D is very similar to Pattern C, with one exception: it iterates back through the Broaden–Narrow–Display cycle (indicated in the figure with curly brackets). These 14 searches generally begin with 1-4 Add Concept moves after the Initial Concept move, followed by at least one SERP Click move and, in some of the searches, some Browse moves. The displays are followed by 1-4 Delete Concept moves, followed in the same query by 1-2 Add Concept moves, an Add Term move, or a Narrow Term move. In most cases, some results are then displayed, and the broadening – narrowing cycle is repeated. In this group of strategies, there is more variability than in Patterns A-C, and some of the searches also incorporated one or more Replace Term moves.

The 18 searches using Pattern E do not include any moves to broaden the searches, but they all include moves to replace terms for concepts already included in the search. Pattern E is Narrow–Display–{(Narrow)–Replace–Display}. The square brackets indicate moves that occur in only some of the searches; the curly brackets indicate a section of the pattern that is repeated. These searches begin with 1-2 Add Concept moves after the Initial Concept move, followed by a results display (mostly SERP Click moves). About one-third of the searches are then narrowed further, using 1-2 Add Concept moves, an Add Term move, and/or a Narrow Term move. In all the searches, the existing concepts are next modified with one or more Replace Term moves. Over one-third of the searches go through at least one more cycle of Replace and Display moves. They conclude with display moves, including both SERP Click and Browse moves.

Pattern F is Narrow–Display–Broaden–Narrow–Display–{Replace–Display}. In this pattern, the Replace Term move makes a more frequent and consistent appearance. These 12 searches begin with 1-2 Add Concept moves following the Initial Concept move. Next, the results are examined with SERP Click moves and, for several of them, multiple Browse moves. Then 1-2 Delete Concept moves or a Delete Term move are used to broaden the search. In the same query, searches are then narrowed through the use of 1-2 Add Concept moves or a Narrow Term move. After viewing the SERP and sometimes clicking on a result, the searches are modified with a Replace Term move. Five of the searches repeat the Replace – Display cycle as many as three more times before concluding. Those that conclude without repeating this cycle usually end in a Browse move; those that repeat the cycle usually end with a SERP Click move.

Pattern G is Narrow–Display–Narrow–(Replace)–Display–Broaden–(Replace/Narrow)–Display. After the Initial Concept move, these 10 searches use 0-3 Add Concept moves. In the following display cycle, the focus is on SERP Click moves, with the majority of searches using this move for examining the results. The search is then narrowed further, using 1-2 Add Concept moves, an Add Term move, or a Narrow Term move. One-third of these searches next include a Replace Term move, and all then go on to a display involving 1-2 SERP Click moves. Next, 1-3 Delete Concept moves, 1-2 Delete Term moves, or a Broaden Term move are used to broaden the search. This is followed by a Replace Term move or one of the three moves that can be used to narrow the search. Five of the searches were modified again before concluding with one of the display moves. As is visible in Fig. 2, there is less consistency in this pattern, particularly in the later queries.

While each of these patterns encompasses some variability across individual searches, they are distinct enough to be recognizable as patterns. All of the patterns include moves that will narrow the search and most include moves that will broaden the search. Three of the patterns include moves to revise the search by replacing the terms being used for a particular concept, changing its meaning but not necessarily increasing or decreasing the results set. All of the patterns include instances of both SERP Click moves and Browse moves, most often paired, with Browse moves following SERP Click moves. Notably, this pattern did not hold up in Groups E, F, and G, where SERP Click moves often were used to view the results with no further browsing. Viewing a SERP without clicking on any of the results occurred primarily in Patterns A and B. This move occurred occasionally in the other groups, but not consistently.

4.3.1. Impact of task on search strategy

One of the goals of this study was to explore the ways in which a particular assigned task might have an impact on the search strategies developed by the study participants. This goal was addressed by focusing on the tasks, to see which patterns were dominant among those searches conducted in response to each task. Because only 114 searches (forming 7 search patterns) were distributed among 20 tasks, no tests of statistical significance could be meaningfully conducted; however, a cross-tabulation of assigned task by search strategy pattern (including single-query searches and very long searches) yielded some preliminary findings. Several tasks elicited a variety of search strategy patterns with no one pattern dominating; in particular, the Analyze (level 3) task in the Health domain and the Analyze (level 3) and Create (level 5) tasks in the Science & Technology domain elicited searches in 5 of the 7 search strategy pattern sets, plus additional single-query searches. In spite of this lack of relationship between task and search strategy pattern, this analysis did yield some preliminary evidence that the assigned search task can have an impact on the search strategies elicited.

The Remember (level 1) tasks across all domains were dominated by single-query searches: 7 of the 11 searches in the Commerce domain, 7 of the 12 searches in the Entertainment domain, 8 of the 12 searches in the Health domain, and 10 of the 12 searches in the Science & Technology domain consisted of only a single query. These tasks can be characterized as simple look-up tasks, even though they included 3-5 facets. As noted earlier, the single-query searches were not analyzed further, since they did not include a reformulation of the search strategy. Even so, we can conclude that tasks at this level of cognitive complexity are likely to yield single-query searches, even if they incorporate as many as five facets in the search task description.

The Understand (level 2) task in the Commerce domain incorporated three separate facets: coffee, fair trade, and labelling requirements (see Table 1, above). Seven of the 12 searches responding to this task used either Pattern A (Narrow–Display–Narrow–Display) or Pattern B (Narrow–Display–Broaden–Display). Among the three searches using Pattern A, most of the narrowing moves were Add Concept moves, with just a few Add Term moves used. SERP Clicks were used must more frequently than Browse moves, but all three searches included at least one Browse move. With one exception (a Narrow Term move), the four searches using Pattern B used only Add Concept moves to narrow the search and used only Delete Concept moves to broaden the search. All four used a mix of both SERP Click and Browse display moves. In summary, this task elicited search strategies that were relatively simple but still required some reformulation for success.

The Evaluate (level 4) task in the Science & Technology domain focused on the potential impact of social media on communication skills (see Table 1, above). It was most often addressed with Pattern B (Narrow–Display–Broaden–Display); four of the 12 searches conducted in response to the Science & Technology/Evaluate task used this search strategy pattern. These four searches each began with an Initial Concept move followed by two Add Concept moves (plus one included an Add Term move), followed by SERP Clicks (plus one included a Browse move). The moves used to Broaden the search were primarily the Broaden Term move (plus one Delete Concept move and one Delete Term move). Three of the four searches elicited by the communication skills task used only the Broaden Term move to broaden the search. This heavy use of the Broaden Term move is of particular interest, since it was used relatively infrequently in the entire data set, occurring only 21 times across all 114 searches (1698 moves); in the searches for the communication skills task, Broaden Term occurred four times (i.e., almost 20% of the Broaden Term moves occurred in these three searches). While a strong quantitative trend emerged in this analysis, a closer look at the specific terms that were broadened in each search did not clarify why this task elicited such a high proportion of Broaden Term moves.

The Analyze (level 3) task in the Entertainment domain focused on the risks to amateurs of trying out extreme sports (see Table 1, above). It primarily elicited searches using Pattern E (Narrow–Display–Broaden–Narrow–(Display)–Edit–Display). Five of the twelve searches conducted on this task fit this pattern; looking at it another way, five of the eight searches using this pattern were elicited by this one task. Several of these five searches were very similar, first searching for a list of extreme sports and then searching for the risks associated with each of several specific sports. For example, one searcher began with the query: *list of extreme sports*. After viewing a couple of web pages that list sports, the searcher entered a series of specific queries (*risks associated with bungee jumping; risks associated with sky diving; risks associated with scuba diving; risks associated with snorkeling; risks associated with indoor climbing*; and *risks associated with surfing*), viewing a single web page from each results list. It is not surprising that this task, by asking for "different types" of extreme sports that might be considered, would elicit this search strategy pattern.

The Create (level 5) task in the Commerce domain focused on shopping for living room furniture. It elicited similar iterations of the Replace Term move. Only one search on this task was included in pattern Group E, but six of the 11 were excluded from further analysis because they were extremely long searches. Among those six long searches, sequences of two to six consecutive Replace Term moves were used. While it does not ask for types of living room furniture, this task does explicitly list several specific items: "you want to get a new couch, chair, rug, television and lamp." The living room furniture shopping task, thus, was very likely to elicit a search strategy incorporating iterations of the Replace Term move.

The remainder of the extremely long searches were elicited by only two of the tasks: the Evaluate (level 4) task in the Commerce domain (purchasing one of three SUV models), and the Create (level 5) task in the Entertainment domain (suggesting a mascot for a new Triple-A baseball team). These searches were too long and varied to categorize, but it can be concluded that tasks of higher cognitive complexity are likely to elicit long searches (in terms of number of moves). Across the searches conducted for the three tasks that elicited very long searches, they averaged almost 20 query moves per search plus 20 display moves per search; these searches ranged from 11-37 query moves per search and 11-31 display moves per search.

5. Discussion

The purpose of this analysis was to evaluate an existing set of 20 search tasks in terms of the search behaviors (i.e., query and display moves) they elicited. More specifically, it addressed the question of whether search tasks can be designed to elicit particular types of search behaviors, or at least increase the likelihood that such behaviors will occur. Within the larger context of designing user studies in the field of interactive information retrieval, it was hoped that a close examination of the search moves used by study participants, across search tasks that exemplify a wide range of levels of cognitive complexity in a number of domains, could provide guidance for researchers as they design future studies.

The search tasks in this study varied by domain and level of cognitive complexity. The four domains were commerce, entertainment, health, and science and technology. The five levels of cognitive complexity were based on Anderson & Krathwohl's (2001) taxonomy of educational objectives; the levels used in this study were Remember, Understand, Analyze, Evaluate, and Create. We found that different search behaviors were associated with different levels of cognitive complexity and different domains. Specifically, more complex tasks elicited more search moves per search; in particular, the Create (level 5) tasks elicited more moves than the other levels. Across the four domains, the Commerce tasks elicited more search moves per search. These findings will be discussed below, in terms of the overall frequency of each type of move and the search strategy patterns in which they were included, the impact of cognitive complexity on the moves and strategies used, the impact of domain on the moves and strategies used, and the ways in which researchers might design search tasks in order to elicit particular types of moves or strategies.

5.1. Search moves/strategies used

In completing the 235 searches assigned to them, study participants made 3101 moves. Of these 1282 (42%) were query moves, and 1819 (58%) were display moves. The two most frequently-used query moves were Add Concept and Delete Concept, accounting for 62% of all the query moves. It should be noted that these two moves both focused on changes in the concepts incorporated in the searches, rather than just changing terms for representing the same concept. Adding in the 235 Initial Concept moves, we can see that 80% of the query moves focused on manipulating concepts, while only 20% incorporated alternative terms to express a particular concept. This finding is consistent with early studies of search moves. For example, in Fidel's (1991) study of professional searchers, only 4.5% of the 1246 moves in MedLine were to broaden a descriptor (i.e., replace one thesaurus term with a broader thesaurus term) and only 2.0% were to find a "better" descriptor (i.e., replace one thesaurus term with a different thesaurus term). The dominance of manipulating concepts rather than terms is also seen in early studies' findings of the dominance of use of the AND operator (to combine concepts) and low use of the OR operator (to combine terms representing a single concept). For example, Shaw & Czaja (1992) found that the OR operator was used in only 2.0% of the 155 searches conducted on a cancer-related database, while the AND operator appeared in 14.7% of the searches. It appears that searchers' inclinations to manipulate concepts, rather than alternative terms that represent a single concept, have endured from the early days of online searching. The recent development of query suggestion tools that display alternative terms for expressing concepts may have an impact on these behaviors, so future studies should continue to examine this question.

Among the display moves, the SERP Click and Browse moves were used approximately the same number of times (854 and 831, respectively). A smaller number of SERP No Click moves (134) occurred. This move code was used to indicate those times when no clicks were made on the SERP before further modifying the query. There are several possible explanations for these findings. One is that the snippets provided on a SERP are sometimes sufficient for sense making and evaluation of the results. Another is that participants were using the SERP to identify additional or alternative query terms, rather than evaluating them as sets of results. A third is that participants are using an examination of the SERP to verify that they believe they have completed the search tasks, i.e., opening additional results will not be fruitful. Future studies incorporating think-aloud protocols could evaluate these alternative explanations, as well as other possibilities.

The study participants composed search strategies from the individual moves. Interestingly, a small number of strategy patterns (7) covered the 114 searches conducted and analyzed qualitatively. Given the wide variability in many search behaviors, it is somewhat surprising that there was so much similarity in the search strategies used. The most frequently-used patterns were A, Narrow–Display–Narrow–Display (25 of the 114 searches analyzed), and C, Narrow–Display–Broaden–Narrow–Display (22 searches). These two patterns might be considered basic patterns, since they are only slightly different from each other and account for such a high proportion (almost half) of the searches.

The most frequently used individual moves were found in a high proportion of the search strategy patterns. The Add Concept move was found in all seven patterns (114 searches) and the Delete Concept move was found in five of the patterns (71 searches). Both the SERP Click and the Browse display moves appeared in six of the seven patterns (104 searches); Browse was not included in pattern G. The SERP No Click move appeared only in patterns A and B (38 searches). Thus, the distribution of particular types of moves across the various search strategy patterns did not diverge from what might be expected.

5.2. Impact of cognitive complexity on search moves/strategies used

As noted earlier, the search tasks were at five different levels of cognitive complexity. In general, search tasks at higher levels of cognitive complexity elicited more moves. There was an increasing trend across the five levels of complexity, except that levels 3 and 4 (Analyze and Evaluate) were essentially equal. In addition, there was a much more pronounced increase at level 5 (Create), which generated almost twice as many moves per search as the Evaluate (level 4) tasks. Much of that difference was due to more Browsing moves than expected and more variety in query reformulation, with numerous uses of Narrow Search with Term Manipulation and Delete Concept moves (see Table 7, above). In addition, as the cognitive complexity of the search tasks increased, the proportion of single-query searches decreased and the proportion of multiple-query searches increased. As noted earlier, 107 of the 235 searches involved only a single query. Of these, 60 were for Remember (level 1) tasks or Understand (level 2) tasks, 38 were for Analyze (level 3) tasks or Evaluate (level 4) tasks, and only nine were for Create (level 5) tasks. From these results, we can conclude that those studies that want to examine longer or multiple search interactions should incorporate search tasks at higher levels of complexity in order to elicit a variety of moderately-complex search moves; those studies that are oriented toward single-query searches should incorporate search tasks at the Remember and Understand levels of complexity.

Examination of the search strategy patterns would also encourage the use of search tasks at higher levels of cognitive complexity. The Understand (level 2) and Analyze (level 3) tasks used in this study each generated searches in all seven of the search patterns identified. Patterns A and B were more frequently used to address Understand (level 2) tasks and patterns C and F were more frequently used to address Analyze (level 3) tasks, but all patterns were represented in the multiple-query searches for these tasks. The Evaluate (level 4) and Create (level 5) tasks also elicited a variety of search strategies. The Evaluate (level 4) tasks elicited all the patterns except G, and the Create (level 5) tasks elicited all the patterns except F. Thus, this analysis also suggests that researchers who want to examine longer or multiple search interactions will benefit from assigning search tasks at higher levels of cognitive complexity.

For the qualitative analysis, 14 extremely long searches (averaging 43 moves in just under 10 queries per search) were eliminated from the analysis because of their length and complexity. It is noteworthy that all of the extremely long searches were elicited by only three of the search tasks, all at the highest levels of complexity. Four the searches were elicited by the Evaluate (level 4) task in the Commerce domain (buying an SUV), six of the searches were elicited by the Create (level 5) task in the Commerce domain (buying living room furniture), and four of the searches were elicited by the Create (level 5) task in the Entertainment domain (naming the mascot for a new baseball team). The types of moves used in these extremely long searches were similar to those used in the entire set of 235 searches, with a few exceptions. These searches included a higher proportion of Delete Concept moves (13.6% as compared to 5.9% in the complete set of searches) and Replace Term moves (8.7% as compared to 4.4%), and a lower proportion of Initial Concept moves (2.3% as compared to 7.6%), as would be expected. Thus, we can conclude that, in these extremely long searches, the study participants reformulated their searches more often than in searches of a more typical length.

In past studies, the number of facets in the task description was used to operationalize task complexity (Wildemuth et al., 2014). Therefore, we analyzed the relationship between number of facets and the search behaviors found in this study. The number of facets in the task description was weakly related to the total number of moves per search (Kendall's tau = 0.246, $p < 0.01$) and to the number of query moves per search (Kendall's tau = 0.165, $p < 0.01$). All 14 of the very long searches were for tasks with 6 or 7 facets. Most (76 of 107) of the single-query searches were conducted for tasks with 4 or 5 facets, but 10 were conducted for tasks with 6-9 facets. From these findings, we would conclude that differentiation of tasks by their cognitive complexity is a more fruitful path for research than using number of facets as the definition of task complexity. However, researchers will want to consider the number of facets in their task descriptions, as well as the level of cognitive complexity.

5.3. Impact of domain on search moves/strategies used

Each study participant conducted searches in a particular domain (Commerce, Entertainment, Health, or Science & Technology); 55[2] or 60 searches were conducted in each domain. A chi square analysis revealed that there was a statistically significant relationship between domain and the frequency with which particular moves were used (Table 8, above). The searches conducted for Commerce search tasks used less Add Concept moves, less SERP Click moves, and more Browse moves than expected. The Entertainment searches used more Narrow Search with Term Manipulation moves and less Browse moves than expected. The Health searches used less Delete Concept moves than expected. The Science & Technology searches used more Add Concept moves, more SERP Click moves, and less Browse moves than expected. While these differences were detected in our statistical analyses, they are not readily explained by examining the specific search tasks assigned.

The 114 searches analyzed qualitatively (i.e., those of moderate length and complexity) were distributed across domains: 31 for Commerce tasks, 25 for Entertainment tasks, 29 for Health tasks, and 29 for Science & Technology tasks.[3] This distribution is well-balanced, with the exception of more of these searches being conducted in the Commerce domain. Since the Commerce searches were distributed across all seven of the strategy patterns discovered, it seems unlikely that the domain explains this higher frequency (or its inverse: a lower frequency of single-query searches).

While all of the search strategy patterns were represented in one or more searches from each domain, it was notable that pattern D (Narrow–Display – {Broaden–Narrow–Display}) was dominated by searches from only two domains. Six of the 14 searches were in the Commerce domain, and five more were in the Health domain. It is not clear why these two domains so frequently elicited this particular (iterative) strategy; its simpler, non-iterative version, pattern C, was elicited by search tasks in all four domains.

5.4. Impact of particular tasks on search strategies used

It may be more useful to examine particular tasks (i.e., tasks at a particular cognitive complexity level in a particular domain), to consider the search patterns that were elicited and the characteristics of those particular tasks that might have led to particular search patterns. First, we will examine the Analyze (level 3) task in the Entertainment domain; it asked the participants to identify some extreme sports that their sister might try out for her 25th birthday, and the risks associated with each (see Table 1, above). Of the nine searches on this task that involved more than a single query, four of them used pattern C (Narrow–Display–Broaden–Narrow–Display) and five of them used pattern F (Narrow–Display–Broaden–Narrow–Display–{Replace–Display}). Most of the nine searches began with some version of the query, "amateur extreme sports." A few also incorporated the term "list" in their initial query. In both cases, it appears that the searchers were looking for ideas about which extreme sports to consider. Most of the searches then followed up with a query that contained the name of a particular sport, e.g., mountain climbing or sky diving, as well as a term related to the risks of the sport. The searches using pattern F typically went through four to six of these iterations before completing the search. It appears that the structure of the task, asking the searcher to generate a list of possibilities and then examine a particular facet of each, regularly elicited this particular search strategy.

The Remember (level 1) task in the Entertainment domain asked the searcher to find the name of the latest album recorded by a particular band (Wolf Parade). Most of the searches elicited by this task were single-query searches; the five multiple-query searches that were conducted in response to this task took only pattern A (three searches; Narrow–Display–Narrow–Display) and pattern E (two searches; Narrow–Display–{(Narrow)–Replace–Display}). While they used two different patterns, all these searches began with the name of the band and, usually, a term for album. In their later queries, they most often added terms related to the recency of the album release (e.g., latest album or 2012). These more complex searches were very similar to the single-query searches elicited by this task, either entering only the name of the band and browsing from there, or entering all three facets (band name, album, and recency) in the initial query. In general, it might be expected that multiple-query searches at the lowest level of complexity might elicit something similar to pattern A, based on the overall frequencies of the Add Concept move.

The Evaluation (level 4) task in the Health domain asked the searcher to identify and compare the effectiveness of current methods of tattoo removal. While most of the searches conducted for this task were single-query searches, the four that involved multiple queries used only two different patterns: pattern A (two searches; Narrow–Display–Narrow–Display) and pattern F (two searches; Narrow–Display–Broaden–Narrow–Display–{Replace–Display}). By juxtaposing these searches against those discussed in the previous two examples, we can see some similarities. Some of the searches (of pattern A) and the single-query searches were similar to the results for the Remember (level 1) task in the Entertainment domain. Only very basic searches were conducted, and they primarily included multiple facets of the search task in the initial query. Those that used pattern F followed a structure that was similar to the more complex searches conducted for the task related to extreme sports risks, just discussed. The first query usually included a general set of terms to identify various methods of tattoo removal, followed by multiple queries, each on a particular method of tattoo removal. The structure of this task, asking the searcher to generate a list of possibilities and then compare those possibilities, is similar to the structure of the extreme sports tasks and so elicited similar search patterns.

For some studies, the researcher may want to assign search tasks that elicit a variety of search strategy patterns. Three of the tasks used in the current study achieved this type of goal. The Analyze (level 3) task in the Commerce domain was one of these. It asked the searcher to consider what types of personal information were needed for identity theft and how one might protect themselves. Nine of

[2] One less participant completed the Commerce search tasks.

[3] If the 14 extremely-long searches are included, there are 41 for Commerce tasks and 29 for Entertainment tasks.

the 12 searches conducted for this task involved multiple queries. Four different search strategy patterns were used in these nine searches. The Analyze (level 3) task in the Health domain asked searchers to analyze multiple methods of artificial tanning in terms of the health risks associated with each. Ten of the 12 searches conducted for this task involved multiple queries. Five different search strategy patterns were used. The Create (level 5) task in the Health domain asked searchers to develop two 30-minute low-impact exercise routines for an elderly great-grandmother (see Table 1, above). Nine of the 12 searches conducted for this task involved multiple queries, and used four different search strategy patterns. It is likely that the relatively high level of cognitive complexity and the overall structure of the tasks influenced the amount of variety in search strategies used to address them.

For some studies, the researcher may want to elicit single-query or other simple search strategies. Many of the Remember (level 1) and Understand (level 2) tasks could serve this purpose. Three of the tasks used in the current study proved particularly useful for such a goal. The Remember (level 1) task in the Science & Technology domain, which asked for the name of the deepest point in the ocean, elicited only two multiple-query searches (the rest were single-query searches). Both of those searches used strategy pattern E (Narrow–Display–{(Narrow)–Replace–Display}). The Understand (level 2) task in the Health domain, which asked about the long-term health risks faced by football players, elicited just two multiple-query searches, both of which used strategy pattern B (Narrow–Display–Broaden–Display). The Evaluate (level 4) task in the Entertainment domain, which asked the searchers to evaluate the evidence for and against the effects of violent video games on teenagers, also elicited just two multiple-query searches, both of which used strategy pattern A (Narrow–Display–Narrow–Display). These examples demonstrate that a variety of types of search tasks may be used to elicit very simple search strategies.

5.5. Additional implications for designing search tasks

The set of tasks evaluated in this study was developed with the hope that they could be re-used by other researchers, as part of a “shared infrastructure” for conducting interactive IR research (Kelly et al., 2015, p.109). In general, this study and prior use of these tasks indicate that they are useful for such purposes. While the set, as a whole, will be a useful addition to researchers’ toolkits, a few of the tasks were found to have weaknesses and should be revised before they are used again. In particular, the Evaluate (level 4) task in the Science & Technology domain (see Table 1, above) asks about whether use of computer-mediated communication services have an impact on people's face-to-face communication skills. This task description uses several key terms that are ambiguous (e.g., the term, communication, might be interpreted in either a technical or a social sense). This type of ambiguity, while realistic, caused difficulties in coding the search moves and, thus, in interpreting the search strategies used. Therefore, as researchers use and adapt these search tasks to their own purposes, we hope that they will share their experiences in applying them, as well as the results of their studies.

As suggested in our earlier discussion of the impact of cognitive complexity on search behaviors, we would encourage researchers to re-use the Create (level 5) tasks and also to develop additional Create tasks in other domains. The fact that these tasks elicited longer and more varied search strategies was not surprising, since there are various attributes of creating, as a task goal, that would lead to such search behaviors. First, the cognitive activity of creating is consistent with Campbell's (1988) idea that complexity may involve multiple paths to the desired end state, multiple desired outcomes to be achieved, conflicting interdependence among the paths to multiple outcomes, and uncertain or probabilistic links among paths and outcome. When a study participant is asked to come up with a mascot for a new baseball team or to create a workout program for their great-grandmother, all of these aspects of complexity are at play. Tasks like these require divergent thinking from study participants, and so will be of particular interest in studies of exploratory search. In addition, they can be usefully incorporated into studies of information use and information creation – new research frontiers for information behavior researchers. While these types of tasks have many positive aspects, researchers will also need to keep in mind that the outcomes from such tasks have no right or wrong answers, so searcher performance cannot be evaluated in traditional ways. Such studies will necessarily be focused on process, rather than a particular end point. Kelly et al., (2015) linked the different cognitive complexity levels (e.g., remember, understand, create) with target outcomes (e.g., fact, list, plan) and mental activities (e.g., identify, compile, describe, compare, decide), which provides another way to use cognitive complexity to design search tasks.

The current study, as expected, found little evidence that domain had a strong influence on search behaviors. The tasks in the Commerce domain elicited longer searches, but there were no discernible differences in the searches elicited across the other three domains. However, researchers will want to re-examine this question, taking into account the possibility that work tasks within a particular domain are patterned by the practices within that domain (Taylor, 1991). One example of this connection is the Create (level 5) tasks in the Commerce domain, which asked participants to shop for a new living room “couch, chair, rug, television and lamp” that match and cost no more than a total of $2500. The task structure might be described as a shopping list of items (i.e., living room furniture) that can be searched for as a group or as individual items. The searches elicited by this task often involved the replacement of query terms, to search for each of the items on the list. While we did not anticipate that the searches elicited by this task would have such consistent patterns of moves, in hindsight it is not surprising that this task structure (and accompanying search strategy pattern) was included among the tasks in this domain. For researchers investigating the search behaviors in domains related to particular professions, especially when specialized databases are used, rather than the everyday life domains developed in this task set, the link between domain and task structure is likely to be even stronger and should be taken into account as task descriptions are developed. It is also the case that, for everyday life domains, a person's interest in a task is likely to vary by domain, and this variation may or may not impact search behaviors.

Finally, the patterns of search strategies identified in this study (Fig. 2, above) may usefully serve as reference models for the construction of new tasks. Researchers typically design tasks around a task type or domain, and do so using a top-down approach.

However, starting instead with patterns of search moves might allow researchers to conceptualize and engage in task construction in novel ways, in particular by inferring task structures from search patterns. For example, pattern F was often used for tasks that asked participants to generate a list of possibilities and then examine a particular facet of each. For example, identifying risks associated with specific extreme sports was a task used in this study; another task with the same structure would be gathering application deadlines for a set of colleges. Another pattern that might be used to create tasks is pattern A, where participants primarily interacted with SERPs rather than browsing through pages. Researchers interested in evaluating or comparing SERP features might design search tasks around this pattern to maximize the amount of time a person spends on the SERP and the person's exposure to the SERP features being studied. Thus, using search move patterns as reference models for task construction has the potential to improve the validity of a variety of interactive IR studies.

6. Conclusion

The ways in which people formulate their online searches, and the ways in which they reformulate those searches when not completely successful, are of interest to a number of constituencies, including those who design information retrieval systems and those who educate searchers on more effective approaches to search strategy formulation. While historically, this topic has been investigated by a number of researchers, it has not been investigated from the perspective of search task development for controlled laboratory studies. In such studies, search tasks are typically assigned to searchers; the approaches people take to address the tasks can vary and sometimes be counterproductive to the researcher's goals. For example, a researcher who is interested in evaluating a novel search interface to support exploratory search would likely be disappointed if a research participant issued a single query, navigated to a Wikipedia page and then declared their search complete. When designing search tasks, it is often difficult to know *a priori* what types of search behaviors the task will elicit even if the task is piloted.

In this paper, we analyzed the sequences of individual search moves made by study participants completing search tasks specifically designed to elicit different types of search strategies from searchers, including those involving more sustained interactions with the system. We found that the cognitive complexity of the tasks and the domain of the tasks both affected the search moves used to complete the tasks; in addition, a small number of search move patterns accounted for all the searches used. This analysis supplements the results presented in Kelly et al., (2015), who reported differences only at the task level and did not look closely at individual search moves. Future investigations of the moves used to address other search tasks can yield benefits for the interactive IR research community. Results from such studies will guide researchers in choosing or constructing tasks that will best fit their study goals. In addition, deeper understanding of the ways that people formulate and reformulate their queries may inspire new approaches to query suggestion and other user support approaches.

Acknowledgments

We thank Dr. Jaime Arguello, Dr. Wan-ching Wu and Dr. Ashlee Edwards, all from the School of Information and Library Science at UNC, for their help collecting and preparing the data. In addition, we thank Dr. Nicholas Belkin and other members of the New Jersey Chapter of the Association for Information Science & Technology for their helpful feedback on a preliminary version of the findings.

References

Agresti, A. (2002). *Categorical Data Analysis* (2nd edition). Hoboken, NJ: John Wiley & Sons, Inc.
Albertson, D., & Meadows, C. III (2011). Situated topic complexity in interactive video retrieval. *Journal of the American Society for Information Science & Technology, 62*(9), 1676–1695.
Anderson, L. W., & Krathwohl, D. A. A. (2001). *Taxonomy for Learning, Teaching and Assessing: A Revision of Bloom's Taxonomy Of Educational Objectives,* Longman, New York.
Bates, M. J. (1979). Information search tactics. *Journal of the American Society for Information Science, 30*(4), 205–214.
Bloom, B. S. (1956). *Taxonomy of Educational Objectives. The classification of educational goals*. New York: Longmans, Green.
Borlund, P. (2016). A study of the use of simulated work task situations in interactive information retrieval evaluations: A meta-evaluation. *Journal of Documentation, 72*(3), 394–413.
Bruza, P., & Dennis, S. (1997). Query reformulation on the internet: Empirical data and the hyperindex search engine. *Proceedings of Computer-Assisted Information Searching on Internet, RIAO '97* (pp. 488–499). .
Byström, K., & Hansen, P. (2005). Conceptual framework for tasks in information studies. *Journal of the American Society for Information Science & Technology, 56*(10), 1050–1061.
Byström, K., & Järvelin, K. (1995). Task complexity affects information seeking and use. *Information Processing & Management, 31*(2), 191–213.
Campbell, D. J. (1988). Task complexity: A review and analysis. *Academy of Management Review, 13*, 40–52.
Fidel, R. (1985). Moves in online searching. *Online Review, 9*(1), 61–74.
Fidel, R. (1991). Searchers' selection of search keys: III. Searching styles. *Journal of the American Society for Information Science, 42*(7), 515–527.
Hassan, A., White, R. W., Dumais, S. T., & Wang, Y.-M. (2014). Struggling or exploring? Disambiguating long search sessions. WSDM '14. *Proceedings of the 7th ACM International Conference on Web Search and Data Mining* (pp. 53–62). .
Hughes-Morgan, K., & Wilson, M. L. (2012). Information vs interaction – Examining difference interaction models over consistent metadata. *Proceedings of the 4th Symposium on Information Interaction in Context (IIiX)* (pp. 72–81). .
Ingwersen, P., & Järvelin, K. (2005). Information retrieval in context – IriX. *SIGIR Forum, 39*(2), 31–39.
Jansen, B. J., Booth, D., & Smith, B. (2009). Using the taxonomy of cognitive learning to model online searching. *Information Processing & Management, 45*(6), 643–663.
Karimi, S., Scholer, F., Clark, A., & Kharazmi, S. (2011). Domain expert topic familiarity and search behavior. *Proceedings of ACM International Conference on Research and Development in Information Retrieval, SIGIR'11* (pp. 1135–1136). .
Kelly, D. (2009). Methods for evaluating interactive information retrieval systems with users. *Foundations and Trends in Information Retrieval, 3*(1-2), 1–232.
Kelly, D., Arguello, J., Edwards, A., & Wu, W.-C. (2015). Development and evaluation of search tasks for IIR experiments using a cognitive complexity framework.

Proceedings of ACM International Conference on the Theory of Information Retrieval, ICTIR'15 (pp. 101–110). .
Kelly, D., & Sugimoto, C. R. (2013). A systematic review of interactive information retrieval evaluation studies, 1967-2006. *Journal of the American Society for Information Science & Technology, 64*(4), 745–770.
Kim, S., & Soergel, D. (2005). Selecting and measuring task characteristics as independent variables. *Proceedings of the ASIST Annual Meeting. 42*. n.p. Retrieved 8/5/2016 from http://dx.doi.org/10.1002/meet.14504201111.
Kinley, K., Tjondronegoro, D., Partridge, H., & Edwards, S. (2014). Modeling users' web search behavior and their cognitive styles. *Journal of the Association for Information Science & Technology, 65*(6), 1107–1123.
Lau, T., & Horvitz, E. (1999). Patterns of search: Analyzing and modeling Web query refinement. *Proceedings of the Seventh International Conference on User Modeling, UM99* (pp. 119–128). .
Li, Y. (2009). Exploring the relationships between work task and search task in information search. *Journal of the American Society for Information Science & Technology, 60*(2), 275–291.
Li, Y., & Belkin, N. J. (2008). A faceted approach to conceptualizing tasks in information seeking. *Information Processing & Management, 44*(6), 1822–1837.
Liu, J., Cole, M., Liu, C., Bierig, R., Gwizdka, J., Belkin, N., et al. (2010). Search behaviors in different task types. *Proceedings of the 10th Annual Joint Conference on Digital Libraries, JCDL '10* (pp. 69–78). .
Liu, J., Kim, C. S., & Creel, C. (2015). Exploring search task difficulty reasons in different task types and user knowledge groups. *Information Processing & Management, 51*, 273–285.
Monchaux, S., Amadieu, F., Chevalier, A., & Mariné, C. (2015). Query strategies during information searching: Effects of prior domain knowledge and complexity of the information problems to be solved. *Information Processing & Management, 51*(5), 557–569.
Pharo, N., & Krahn, A. (2011). The effect of task type on preferred element types in an XML-based retrieval system. *Journal of the American Society for Information Science & Technology, 62*(9), 1717–1726.
Rieh, S. Y., & Xie, H. (2006). Analysis of multiple query reformulations on the Web: The interactive information retrieval context. *Information Processing and Management, 42*, 751–768.
Sharpe, D. (2015). Your chi-square test is statistically significant: Now what? *Practical Assessment, Research & Evaluation, 20*, 8. Available online: http://pareonline.net/getvn.asp?v = 20&n = 8.
Shaw, D. J., & Czaja, R. F. (1992). User interactions with the PDQ cancer information system. *Bulletin of the Medical Library Association, 80*(1), 29–35.
Shiri, A. A., & Revie, C. (2003). The effects of topic complexity and familiarity on cognitive and physical moves in a thesaurus-enhanced search environment. *Journal of Information Science, 29*(6), 517–526.
Shute, S. J., & Smith, P. J. (1993). Knowledge-based search tactics. *Information Processing & Management, 29*(1), 29–45.
Taylor, R. S. (1991). Information use environments. In B. Dervin, & M. J. Voigt (Vol. Eds.), *Processing in communication Sciences: volume X*, (pp. 217–255). Norwood, NJ: Ablex.
Thatcher, A. (2006). Information seeking behaviorscognitive strategies in different search tasks on the WWW. *International Journal of Industrial Ergonomics, 36*(12), 1055–1068.
Toms, E., Freund, L., Kopak, R., & Bartlett, J. C. (2003). The effect of task domain on search. *Proceedings of Conference of the Centre for Advanced Studies on Collaborative Research, CASCON '03* (pp. 303–312). . Retrieved December 1, 2017, from https://dl.acm.org/citation.cfm?id = 961370.
Vibert, N., Ros, C., Le Bigot, L., Ramond, M., Gatefin, J., & Rouet, J.-F. (2009). Effects of domain knowledge on reference search with the PubMed database: An experimental study. *Journal of the American Society for Information Science & Technology, 60*(7), 1423–1447.
Wacholder, N. (2011). Interactive query formulation. *Annual Review of Information Science & Technology, 45*, 157–196.
Wildemuth, B. M. (2004). The effects of domain knowledge on search tactic formulation. *Journal of the American Society for Information Science & Technology, 55*(3), 246–258.
Wildemuth, B. M., Freund, L., & Toms, E. G. (2014). Untangling search task complexity and difficulty in the context of interactive information retrieval studies. *Journal of Documentation, 70*(6), 1118–1140.
Wildemuth, B. M., & Moore, M. E. (1995). End-user search behaviors and their relationship to search effectiveness. *Bulletin of the Medical Library Association, 83*(3), 294–304.
Xie, I., & Joo, S. (2010). Transitions in search tactics during the Web-based search process. *Journal of the American Society for Information Science & Technology, 61*(11), 2188–2205.
Xie, I., & Joo, S. (2012). Factors affecting the selection of search tactics: Tasks, knowledge, process, and systems. *Information Processing & Management, 48*(2), 254–270.
Zhang, X., Liu, J., Cole, M., & Belkin, N. (2015). Predicting users' domain knowledge in information retrieval using multiple regression analysis of search behaviors. *Journal of the Association for Information Science & Technology, 66*(5), 980–1000.