The Effects of Search Task Determinability on Search Behavior

Rob Capra, Jaime Arguello, Yinglong Zhang

University of North Carolina at Chapel Hill

Abstract. Among the many task characteristics that influence search behaviors and outcomes, task complexity has received considerable attention. One view of task complexity is through the lens of *a priori* determinability—a measure of how much the searcher knows about the task outcomes, information requirements, and processes involved. In this paper, we explore a novel manipulation of *a priori* determinability in the context of comparative search tasks, which require comparing items (or alternatives) along different dimensions. Our manipulation involved explicitly including the *items* to be compared and/or the *dimension* by which to compare items in the search task description. We report on two user studies that investigate the effects of our manipulation on searchers' pre-task perceptions, search behaviors and post-task outcomes. Our results found that specifying the items had an effect on searchers' pre-task perceptions, but not their search behaviors and outcomes, and that specifying the dimension had no effect on perceptions, but made the task *more* difficult by possibly introducing uncertainty into the search process.

1 Introduction

A large body of prior research has investigated how search tasks vary along different dimensions. Task characteristics can relate to the search task's main activity (e.g., gathering factual information), end goal (e.g., well-defined or amorphous), task structure (e.g., its complexity), or the searcher's perceptions of the task (e.g., its expected difficulty) [11]. Studies have shown that many of these task characteristics can influence search behaviors and outcomes [12, 15]. Understanding how task characteristics influence search behaviors is important to the study and design of interactive IR systems and to the development of models of how users engage in search processes.

Task complexity is one characteristic that has received considerable attention in recent work [2, 3, 7–9]. Task complexity is a multi-faceted concept that has been considered from different perspectives [15]. An influential approach proposed by Byström and Järvelin [5] is to view task complexity in terms of the *a priori* determinability of the task (i.e., how well a searcher is able to determine the outcomes, processes, and information requirements for a task in advance of actually performing it) [5, 13, 3]. A search task with low determinability is one with high *uncertainty* regarding the solution, information requirements, and the processes involved in gathering the needed information.

In this paper, we explore *a priori* determinability as a way to investigate the effects of task complexity on searchers' perceptions, search behaviors, and outcomes. Our goal was to *manipulate* the determinability of tasks while holding other task characteristics constant. To this end, we focus on *comparative* tasks. Our study participants were asked to search for information in order to compare and contrast items (or alternatives) belonging to the same category. For example, one of the tasks asked participants to compare and contrast different methods for purifying water during a hiking trip. Comparative tasks involve two important activities: (1) identifying the different *items* belonging to the given category (e.g., water filters, chemical tablets, boiling techniques) and (2) identifying the different *dimensions* by which the items can differ (e.g., the weight of the equipment, the time it takes to purify the water, the micro-organisms eliminated). We created 17 different task groups with 4 determinability levels each. Our four determinability levels were operationalized by explicitly including or excluding the items and dimensions in the task description.

We report on two crowdsourced studies (Study 1 and Study 2) that investigate the following three main research questions (RQ1-RQ3). In RQ1, we investigate whether searchers perceive differences in determinability when we include items and/or dimensions in a comparative search task description. In RQ2, we consider whether our manipulation of determinability yields differences in search behaviors and strategies. Finally, in RQ3, we investigate whether our manipulation of determinability yields differences in perceived outcome measures (e.g., difficulty, engagement, satisfaction) reported after completing the task. Study 1 investigates RQ1, and Study 2 investigates RQ2 and RQ3.

2 Related Work

Our research builds on prior work focused on understanding how task characteristics influence search behaviors and outcomes.

Tasks play an important role in understanding information seeking and searching [14]. Byström and Hansen [4] distinguish between *work tasks*, *information-seeking tasks*, and *information search tasks.* A search task is done in the context of an information-seeking task and both are done in the context of a work task. In this paper, we manipulate determinability at the information search task level.

A large body of prior work has characterized tasks along different dimensions. Li and Belkin [11] provide an extensive literature review and propose a classification scheme, including aspects of the task's activity, goal, and structure.

Different characterizations of task complexity have been proposed in prior work (see Wildemuth *et al.* [15] for a review). Campbell [6] characterized task complexity in terms of: (1) the number of required outcomes, (2) the number of paths to the outcomes, (3) the level of uncertainty about the paths, and (4) the degree of interdependence between the paths. Jansen *et al.* [8] (and later Kelly *et al.* [9]) used Anderson and Krathwohl's taxonomy of learning outcomes from educational theory [1] to create tasks with different levels of *cognitive* complexity. Cognitive complexity is associated with the amount of learning and mental effort required to complete the task. The simplest tasks (called *remember* tasks) require verifying or searching for a specific fact, while the most complex tasks (called *create* tasks) require searching in order to develop a new solution to a problem.

More closely related to our work, Byström and Järvelin [5] (and later Bell and Ruthven [3]) reduced task complexity to the *a priori* determinability of the task. Byström and Järvelin [5] defined *a priori* determinability as the extent to which a searcher is able to internalize the task at hand and deduce: (1) the task outcomes, (2) the information needed to produce the outcomes, and (3) the processes associated with gathering the required information. In later work, Bell and Ruthven [3] sought to *manipulate* the *a priori* determinability of tasks in a study. Tasks were designed to influence the *a priori* determinability of: (1) the information needed, (2) the strategy for searching, and (3) the need to synthesize information from multiple sources.

Similar to our research, past studies have investigated how different characterizations of task complexity influence participants' expectations, behaviors, and outcomes. Studies have found that complex tasks are associated with higher levels of expected difficulty [9, 7, 3], experienced difficulty [2, 7, 9, 3], and search effort as indicated by measures derived from queries, clicks, bookmarks, and the task completion time [2, 7–9]. Additionally, Kelly *et al.* [9] found that participants' choice of queries, query-terms, and pages visited *diverged* more from each other during complex tasks. Finally, Capra *et al.* [7] found that task complexity affected participants' engagement with a search assistance tool.

Our research adds to this body of work by investigating how a novel manipulation of task determinability in the context of comparative search tasks affects users' expectations, search behaviors, and experiences.

3 Determinability of Comparative Tasks

In this work, we manipulated the determinability of comparative tasks. Our tasks asked participants to compare and contrast items or alternatives belonging to the same category. Comparative tasks fall under the *analyze* level of cognitive complexity. According to Anderson and Krathwohl [1], *analyze* tasks require "breaking materials or concepts into parts and determining how the parts relate to each other", and may involve mental and physical activities such as "organizing and differentiating" and "creating spreadsheets". Comparative tasks involve two important activities: (1) identifying the different items associated with the given topic, and (2) identifying dimensions by which the items can differ.

Our manipulation involved making the task narrower in scope by specifying the items to be compared and/or the dimension by which to compare the items. We created 17 task topics (groups) with 4 determinability levels each, for a total of 68 task descriptions.[1] Each task included a background story that motivated the information need. The background story was consistent across all task descriptions within the same group, and the final information request was manipulated to elicit different levels of determinability. Below, we illustrate our four determinability levels for one task group. The items and dimension are shown in bold.

– **Unspecified (U):** no items or dimension specified.
"You are planning an extended hiking trip. You heard that it can be unsafe

[1] Task descriptions are available at https://ils.unc.edu/searchstructures/resources/ecir2017_tasks.txt.

to drink water directly from streams along the trail and that you need to purify water before drinking it. You would like to learn more about this. For this task, find out: What are different methods for purifying water to drink from streams and how do they differ?"

- **Specified Items (I):** specified two items to compare, but not the dimension. "You are planning... For this task, find out: How do **boiling water** and using a **charcoal filter** differ as methods for purifying water from streams?"
- **Specified Dimension (D):** specified the dimension, but not the items. "You are planning... For this task, find out: What are different methods for purifying water to drink from streams and how do they differ in terms of the **micro-organisms eliminated**?"
- **Both (B):** specified both items and the dimension. "You are planning... For this task, find out: How do **boiling water** and using a **charcoal filter** differ as methods for purifying water from streams in terms of the **micro-organisms eliminated**?"

Figure 1 illustrates a conceptual representation of comparative tasks as a grid to compare items across dimensions. Our *unspecified* tasks (Region U) left the items and dimensions completely open. Our *specified items* and *specified dimension* tasks were more narrowly focused by specifying two items to compare (Region I) or by specifying one dimension by which to compare any number of items (Region D). Finally, our *both* tasks were the most narrowly focused and limited the comparison to two items and one dimension (Region B).

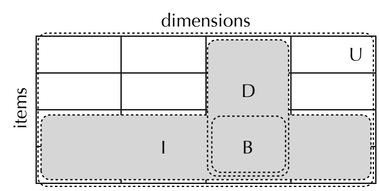

Fig. 1. Conceptual Representation of Comparative Tasks.

Our manipulation of task determinability can also be understood in light of the three factors described in Bell and Ruthven [3]. We expected that making the task more narrowly focused would produce less uncertainly in terms of the: (1) information needed, (2) the strategy for finding relevant content, and (3) the need to consult multiple sources. Table 1 lists the different topics, items, and dimensions associated with our 17 task groups.

4 Study 1: Search Task Evaluation

In our first research question (RQ1), we investigate whether specifying the items and/or the dimension of a comparative task might influence participants' perceptions of the *a priori* determinability and expected difficulty of the task.

To investigate this question, we conducted a crowdsourced study using the Amazon Mechnical Turk (MTurk). Participants were asked to read a series of four search task descriptions and rate their level of agreement with a set of

Table 1. Task topics, items, and dimensions used in our task descriptions.

Topic	Items	Dimension
water purification methods	boiling water and charcoal filter	micro-organisms eliminated
carpal tunnel treatments	wrist splints and corticosteroids	side-effects
medicinal herbs for colds	Echinacea and St. John's Wort	effectiveness
motor oil for cars	synthetic and organic motor oils	performance
types of rice	rice and brown rice	fiber content
types of lightbulbs	fluorescent bulbs and LEDs	type of light
types of ballet	classical and neoclassical	physical movements
music speaker materials	polypropylene and paper	high-frequency performance
garden fertilizers	organic and chemical fertilizers	nutrient content
types of paint thinner	linseed and poppyseed oil	effect on paint color
wifi routers	single band and dual band	signal interference
different types of plastic	PET and PVC	ability to be recycled
indoor dog breeds	Pug and Bichon Frise	amount of exercise needed
smoking cessation methods	nicotine gum and nicotine patches	success rate
covering material for couch	leather and microfiber	ease of cleaning
Chinese keyboards	Pinyin and Wubi methods	difficulty of learn
cooking skillet materials	aluminum and cast iron	ability to distribute heat

10 statements about their impressions of each task. Participants were asked to imagine that they were going to search for information using a web search engine in order to answer the task. Since we were primarily interested in participants' impressions of the task, participants did not actually perform the search. To gain statistical power, we designed Study 1 with task determinability as a within-subject factor. To keep the study manageable, we chose a subset of four task topics (carpal tunnel treatments, motor oil for cars, garden fertilizers, and types of plastic). Each participant did all four tasks, each with a different determinability level. Treatment combinations (n=16) were created using a Latin square and participants were randomly assigned to one. Ultimately, we collected usable data from 63 participants.

Through a series of small-scale pilot tests, we developed a set of 10 statements (Table 2) to measure participants' perceptions about the task descriptions. These statements inquire about a range of concepts related to determinability and expected difficulty, including: prior knowledge, how focused the task is, whether it includes new information previously unknown to the participant, and the scope of the task in terms of the items to be compared and the dimensions by which to compare them. Participants indicated their level of agreement with the 10 statements on a 7-point scale from strongly disagree (1) to strongly agree (7). The statements were displayed below the task description. Participants were not allowed to go back after submitting responses for a task.

We recruited U.S. MTurk workers with a $\geq 95\%$ acceptance rate. Participants were paid $0.75 USD and were allowed to complete the study only once.

Study 1 Results (RQ1): Table 3 summarizes the results for each of the 10 measures for each level of task determinability. We conducted one-way repeated measures ANOVAs to investigate the differences of determinability on each measure. Table 3 shows the results of this analysis, along with post-hoc comparisons using the modified Bonferroni correction outlined in Keppel [10](p.170).

Table 3 shows three important trends. First, our manipulation of determinability had a significant effect for 8 of the 10 measures—only Difficulty and LookFor did not show significant differences. It is possible that participants reported similar levels of expected difficulty because all task descriptions were

Table 2. Study 1 Questionnaire Measures

Measure	Agreement Statement
PriorKnow	I already know a lot about this topic.
Specificity	The task is very specific.
Difficulty	I think the task will be difficult.
Focused	The information requested is narrowly focused.
NewInfo	The task description provides me with information that I did not already know.
LackDim	There are dimensions of the task that are not specified in the description.
ManyDetail	The task description has a lot of details.
LookFor	Right now, I know some specific things to look for to address the task.
SpecItems	The task is very specific in terms of the number of items I need to compare.
SpecDim	The task is very specific in terms of the factors I need to consider when comparing the items.

Table 3. Study 1 Questionnaire Results. The task versions with specified items are shown with gray background

	Unspecified (U)	Items (I)	Dimension (D)	Both (B)	*p*	post-hoc
PriorKnow	3.19 (1.62)	2.47 (1.60)	2.97 (1.70)	2.34 (1.38)	.001	U>I,B; D>B
Specificity	5.07 (1.48)	5.52 (1.27)	5.20 (1.39)	5.72 (1.07)	.000	U<I,B; D<B
Difficulty	3.52 (1.55)	3.61 (1.56)	3.66 (1.63)	3.87 (1.53)	.532	
Focused	4.51 (1.45)	5.19 (1.29)	4.79 (1.32)	4.95 (1.34)	.007	U<I,B
NewInfo	3.59 (1.70)	4.15 (1.83)	3.75 (1.62)	4.18 (1.76)	.027	U<I,B
LackDim	3.82 (1.58)	3.19 (1.50)	3.55 (1.59)	3.44 (1.48)	.057	U>I
ManyDetail	3.88 (1.44)	4.38 (1.39)	4.22 (1.55)	4.52 (1.54)	.032	U<I,B
LookFor	5.08 (1.18)	5.33 (1.17)	5.43 (1.13)	5.08 (1.52)	.131	
SpecItems	4.00 (1.65)	5.15 (1.51)	4.50 (1.76)	5.21 (1.56)	.000	U<D,I,B; D<I,B
SpecDim	4.83 (1.40)	5.02 (1.42)	4.98 (1.35)	5.35 (1.22)	.050	U<B

associated with comparative tasks at the *analyze* level of Anderson and Krathwhol's taxonomy [1].

The second important trend is that specifying the items in the task description had a strong effect on many of our measures. This can be seen by comparing task versions I and B (the two where the items were specified) with task versions U and D (the two where the items were *not* specified). The observed differences were generally in the directions we expected—the tasks with the specified items (I and B) were perceived to be more focused, had more details, and were more specific in terms of the items and dimensions to be considered. Similarly, the tasks with the specified items (I and B) were perceived to provide more information that the participant did not already know (NewInfo), and influenced participants to rate their prior knowledge as being lower (PriorKnow).

The third important trend is that specifying the dimension in the task description did *not* have a strong effect. This can be seen by comparing pairs of tasks where the only difference was the specified dimension (compare U vs. D and I vs. B). Based on our post-hoc comparisons, task versions U and D, as well as task versions I and B, were statistically equal for 9 out of 10 measures. SpecItems was the only measure for which specifying the dimension had a significant difference (U<D).

We were also interested in understanding how participants interpreted our 10 measures. To investigate this, we conducted an exploratory factor analysis with principle components analysis (using varimax rotation) and found a solution using two factors that explained 51% of the variance. All measures had factor loadings $\geq .6$ for these two factors, with no measures having a cross-loading $> .4$. Because the measures had loadings $\geq .6$, we kept them all in our final solution.

The final factor loading matrix is shown in Table 4, and suggests that our questions measured two main concepts. Factor 1 focuses on the expected difficulty of the task. These measures were inversely related—when participants perceived the task as specifying new information, they reported having less prior knowledge, and expected the task to be more difficult. Factor 2 focuses on the determinability of the task—the extent to which the task specified the information needed to complete it and reduced uncertainty about what to look for. Most of the measures loaded on this factor and were directly related. LackDim was negatively weighted because of its negative wording (Table 2). These results also suggest that participants did not make a strong distinction between our questions about the specification of items versus dimensions.

Table 4. Study 1 Factor Analysis

Measure	Factor 1	Factor 2
PriorKnow	-.606	
Specificity		.752
Difficulty	.693	
Focused		.741
NewInfo	.620	
LackDim		-.612
ManyDetail		.636
LookFor		.702
SpecItems		.661
SpecDims		.735

5 Study 2: Search Behaviors and Outcomes

In our remaining research questions, we investigate whether and how specifying items and/or dimensions of a comparative task might influence participants' search behaviors (RQ2) and perceptions about their search experience (RQ3).

To investigate these questions, we conducted a second crowdsourced study on Amazon Mechanical Turk. Participants were given a search task, were asked to search for and bookmark relevant pages, and were asked to complete a post-task questionnaire. Search tasks were presented as individual HITs on MTurk.

Each HIT presented a single task description and asked the participant to find and bookmark pages that would help them construct a response for the task. Searches were conducted using a custom-built search system that produced results using the Bing Web Search API. Participants were required to install toolbar buttons for bookmarking pages, viewing the current set of bookmarks in a pop-up window, and indicating when they were done with the task. When bookmarking a page, participants were required to provide a brief justification for why the page was useful, and the "view bookmarks" pop-up window allowed participants to delete bookmarks. Participants were required to bookmark at least 3 pages before finishing the HIT. Participants were paid \$0.30 USD per HIT and were offered an \$0.30 USD bonus if they bookmarked five or more pages. After finishing the task, participants completed a post-task questionnaire with questions about their level of enjoyment, engagement, interest increase, knowledge increase, perceived task difficulty, satisfaction with their solution and search strategy, and about how much time pressure they felt.

For each of our 17 task groups and 4 levels of task determinability, our goal was to collect data from 15 to 20 participants. To this end, we posted a total of 1,360 HITs on MTurk ($17 \times 4 \times 20$). Participants were randomly assigned to task-group/determinability-level combinations, but were *not* allowed to complete tasks from the same group (topic) more than once. Once all combinations of task topic and determinability level had data from at least 15 participants, we stopped the data collection. In total, we collected data for $1,317$ search tasks and 348 participants. We recruited U.S. MTurk workers with a $\geq$95% acceptance rate.

Study 2 Search Behavior Results (RQ2): For this and the next analysis, we conducted one-way ANOVAs to investigate the differences of determinability on each measure. Results for RQ2 are presented in Table 5. The first eight measures are associated with the level of search effort (e.g., number of queries, clicks, bookmarks, time between the query and the first SERP click (if any), and time to completion in seconds). The next two measures suggest trial-and-error (e.g., number of queries without a click and queries without a bookmark). Finally, the last four measures capture the extent to which participants' searches *diverged* from other participants who completed the same combination of task-group and determinability level. The query log-likelihood measure was computed by first generating a language model from all queries issued by the other participants who completed the same task-group/determinability combination, and then measuring the average log-likelihood of the participant's queries. A lower log-likelihood score indicates that the participant's queries contained language that was not frequently used by the other participants. Similarly, the last three measures are associated with the number of queries, query terms, and clicked URLs that were *not* observed in search sessions from the other participants.

Table 5 shows two important trends. The first main trend is that specifying the dimension had a strong effect on search behavior. This can be seen by comparing task versions D and B (the two versions where the dimension was specified) with tasks versions U and I (the two versions where the dimension was *not* specified). In terms of search effort, task versions D and B had significantly more queries, longer queries, and fewer clicks and bookmarks per query. It also took longer for participants to produce the first SERP click after issuing a query, suggesting that participants had more difficulty identifying relevant results. Task versions D and B also had more evidence of trial-and-error (more queries without clicks and bookmarks), although the differences were not significant. Finally, in terms of search strategy, while completing task versions D and B, participants issued significantly more unique queries (as evidenced by the lower query log-likelihood and greater number of unique queries), and clicked on more unique URLs.

The second important trend is that specifying the items did *not* have a strong effect. This can be seen by comparing between pairs of task versions where the only difference was the specified items (compare U vs. I and D vs. B). Both pairs of task versions were associated with similar amounts of search effort and divergence of search strategy. In fact, our post-hoc comparisons revealed no significant differences between task versions U and I and between D and B.

Table 5. Study 2 Search Behavior Results. The task versions with a specified dimension are shown with gray background

	Unspecified (U)	Items (I)	Dimension (D)	Both (B)	p	post-hoc
Queries	1.91 (1.44)	1.93 (1.48)	2.32 (1.61)	2.35 (1.68)	.000	U,I <D,B
QueryLength	7.42 (4.37)	7.76 (3.34)	9.28 (6.99)	9.47 (5.81)	.000	U,I <D,B
Clicks	5.91 (3.10)	6.06 (3.11)	6.38 (3.49)	6.07 (2.96)	.300	–
ClicksPerQuery	4.04 (2.62)	4.04 (2.32)	3.58 (2.23)	3.50 (2.29)	.003	U,I > D,B
TimeToFirstClick	10.92 (24.42)	8.88 (9.55)	11.55 (22.93)	14.60 (33.38)	.028	I <B
Bookmarks	4.35 (1.10)	4.46 (1.17)	4.45 (1.18)	4.35 (1.15)	.417	
BooksPerQuery	3.12 (1.55)	3.20 (1.61)	2.71 (1.54)	2.68 (1.58)	.000	U,I > D,B
CompletionTime	330.56 (238.13)	353.94 (252.38)	374.75 (292.06)	373.21 (288.97)	.128	–
QueriesWOClicks	0.37 (0.80)	0.32 (0.79)	0.39 (0.77)	0.42 (0.81)	.440	–
QueriesWOBooks	0.50 (0.97)	0.47 (0.93)	0.61 (1.07)	0.61 (1.03)	.146	–
QueryLogLike	-46.65 (28.28)	-46.78 (20.26)	-58.85 (45.37)	-57.81 (36.95)	.000	U,I>D,B
UniqueQueries	1.31 (1.49)	1.42 (1.57)	1.82 (1.71)	1.90 (1.79)	.000	U,I <D,B
UniqueQTerms	0.84 (1.56)	0.83 (1.68)	1.02 (1.70)	0.84 (1.48)	.398	–
UniqueURLs	1.14 (1.49)	1.17 (1.51)	1.60 (1.84)	1.31 (1.47)	.001	U,I <D

Study 2 Post-task Questionnaire Results (RQ3): Table 6 summarizes our post-task questionnaire results. Task determinability had a significant effect on several measures: knowledge increase, perceived difficulty, overall satisfaction, and satisfaction with the search strategy.

The trends in this data largely match the search behavior results reported in Table 5—specifying the dimension often had more impact than specifying the items. Table 6 shows that task versions D and B had lower overall satisfaction and lower satisfaction with the search strategy as compared to versions I and U, and that task version B had higher levels of difficulty than versions U and I. Tasks versions D and B also had lower ratings for enjoyment, engagement, and interest increase, but these differences did not reach statistical significance. Interestingly, knowledge increase was highest for task version I. Overall, these trends are consistent with the search behavior results and illustrate how specifying the dimension increased the effort required.

Table 6. Study 2 Post-Task Questionnaire Results. The task versions with a specified dimension are shown with gray background.

	Unspecified (U)	Items (I)	Dimension (D)	Both (B)	p	post-hoc
Enjoyment	4.69 (1.69)	4.73 (1.66)	4.47 (1.76)	4.44 (1.77)	.062	
Engagement	5.02 (1.65)	5.07 (1.71)	4.89 (1.67)	4.90 (1.74)	.448	
InterestInc	4.59 (1.86)	4.74 (1.73)	4.50 (1.89)	4.51 (1.87)	.297	
KnowledgeInc	5.03 (1.57)	5.36 (1.39)	5.09 (1.47)	5.11 (1.56)	.022	U,D,B<I
Difficulty	2.47 (1.42)	2.51 (1.48)	2.62 (1.47)	2.85 (1.58)	.005	U,I<B
OverallSat	5.64 (1.45)	5.61 (1.51)	5.38 (1.50)	5.27 (1.59)	.004	U,I>D,B
StrategySat	5.73 (1.38)	5.67 (1.40)	5.47 (1.46)	5.46 (1.52)	.029	U>D,B
TimePressure	3.25 (1.94)	3.14 (1.89)	3.30 (1.89)	3.24 (1.91)	.773	

6 Discussion

In this work, we set out to explore a novel method for manipulating the determinability of comparative search tasks. Our results reveal interesting points about how our manipulation of items and dimensions influenced participants' pre-search perceptions of a task, as well as their search behaviors and outcomes.

Study 1: Including the items in the task description influenced participants to perceive the task as being more focused and reduced their uncertainty about what to look for. Including the items also led participants to report that the tasks contained new information and that their prior knowledge of the task domain was lower. The same effects were *not* observed when the dimension was included. This was surprising to us. We expected that adding constraints of either type (items or dimensions) would increase participants' perceptions of determinability, and that there might even be an additive effect.

One possible explanation is that participants did not notice the dimension in the task description as much as they noticed the items. Another explanation is that participants did not perceive the dimension as being as strong of a constraint as the items. The items were specified as concrete noun phrases, while the dimensions were often specified as abstract concepts (e.g., "side-effects", "effectiveness", "performance", "success rate", "difficulty to learn"). These results have implications for how we define and operationalize determinability—it is not sufficient to assume that a task with more constraints is perceived to be more determinable. Based on our observations, constraints are not equal in their influence on determinability prior to working on the task.

Study 2: In Study 2, we found two interesting results: (1) specifying the items did not have an effect, and (2) specifying the dimensions did have an effect, but it was the opposite of what we expected. Based on the results of Study 1, it could be expected that specifying the items would make the task easier in Study 2. However, in Study 2, specifying the items did not yield differences in the search process or outcome measures. This may be because in the conditions where the items were *not* specified, participants were able to engage in satisficing behaviors, by bookmarking the most easily found pages or finding pages containing summaries of items.

The second interesting result from Study 2 is that specifying the dimension led to more difficult search tasks, as evidenced by greater levels of search activity, more divergent search strategies, greater levels of experienced difficulty, and lower levels of knowledge increase and satisfaction. In our initial view, we expected that adding both items and dimensions would reduce uncertainty (increasing determinability) and make the tasks easier to complete. However, adding the dimension constraint made the task more difficult, possibly because its determinability was actually *reduced*.

Task determinability involves uncertainly about different aspects of the task—the task inputs, required outcomes, and processes involved. It is possible that specifying the dimension narrowed the scope of the task and therefore reduced the uncertainly of the task outcome, but *increased* the uncertainty of the search process in different ways.

One possibility is that the dimensions of a comparative task may not be natural query-like concepts. For example, consider our "cooking skillet materials" task in Table 1. The dimension required participants to find information on how cooking materials are able to distribute heat uniformly. The language surrounding a dimension may be unknown or varied, making it more difficult to construct effective queries and identify relevant content. To gain more insight, one of the authors manually coded all queries submitted by our Study 2 participants as either containing at least one item and/or containing at least one dimension. Across all determinability levels, there were 1,441 queries with at least one item and 960 queries with at least one dimension. Indeed, this analysis suggests that it was easier for participants to explicitly search for items versus dimensions.

A second explanation is that many of our dimensions (e.g., "side-effects", "effectiveness", "performance", "success rate", "ease of cleaning", "difficulty to learn") may have introduced subjectivity into the task. Including such dimensions may have required participants to judge the credibility of information or synthesize different opinions.

Based on Bell and Ruthven's factors of determinability [3], including the dimension might have *increased* uncertainty in terms of the strategy for searching and identifying relevant content, as well as the need to integrate information from different sources. Interestingly, our results suggest that participants did not recognize this added complexity from the dimension from simply reading the task description (Study 1).

Summary: Our results provide insights into the complex relationship between task constraints and level of determinability. Our results suggest three important findings. First, task constraints that are perceived as making the task more focused may not yield differences in search behaviors and outcomes. In our case, *omitting* the items might have allowed participants to engage in satisficing behaviors when conducting the search (e.g., limiting the search to items found early on). Second, adding constraints to a task may not necessarily make it easier. In our case, specifying the dimension led to more search effort, possibly by introducing more uncertainty into the search process (e.g., constructing queries, identifying relevant content, and dealing with subjective information). Finally, while adding constraints may make a task harder, this may not be perceived before actually working on the task. This is the classic "you don't know what you don't know" paradox. In our case, participants did not perceive tasks with the dimension as being different than those without. However, the dimension led to more search activity, higher levels of difficulty, and lower levels of satisfaction. It is possible that participants experienced the added uncertainty only after starting the task (not by simply reading the description).

7 Conclusion

In this paper, we sought to create tasks with varying degrees of determinability, defined as the level of *uncertainty* regarding the task inputs, outputs, and processes involved. We focused on a specific task type (comparative tasks) and introduced a method for systematically varying task components (the items to be compared and/or the dimension by which to compare them). By including specific items or the dimension in the task description, we expected to narrow the scope of the task, increasing its determinability, and make it easier to complete.

Our results reveal a more complex situation. In Study 1, participants perceived differences in the tasks based on the items, but not the dimensions, possibly because the dimensions were more subtle in the task description. In Study 2, the items did not have an effect on search behaviors and outcomes (possibly due to satisficing behaviors in the absence of the items) and the dimensions actually made the search task harder. Interestingly, adding the dimension might have made the task less determinable by introducing uncertainly into the search process. A post-hoc analysis suggests that it was easier for participants to query for items than dimensions.

Our results have implications for experimental design, the design of search systems, and for frameworks of information seeking. From an experimental design standpoint, our results illustrate how subtle differences in task descriptions can have significant (and unexpected) influences on perceptions of tasks and on search behaviors. Wildemuth *et al.* [15] called for more research to investigate the impacts of task characteristics. Our results address this call, providing a detailed view of the effects of a specific, systematic manipulation of task determinability. From a system design perspective, our results suggest that providing recommendations or choices of dimensions in an interface (e.g., faceted search) may be especially helpful to users working on comparative tasks. Finally, our results provide additional insights into the role of *a priori* determinability in information seeking.

Acknowledgments. This work was supported in part by NSF grants IIS-1552587 and IIS-1451668. Any opinions, findings, conclusions, and recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the NSF.

References

1. L. W. Anderson and D. R. Krathwohl. *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives.* 2001.
2. J. Arguello. Predicting search task difficulty. In *ECIR*. Springer-Verlag, 2014.
3. D. J. Bell and I. Ruthven. Searchers' assessments of task complexity for web searching. In *ECIR*, pages 57–71. Springer-Verlag, 2004.
4. K. Byström and P. Hansen. Conceptual framework for tasks in information studies. *JASIST*, 56(10):1050–1061, 2005.
5. K. Byström and K. Järvelin. Task complexity affects information seeking and use. *Information Processing & Management*, 31(2):191–213, 1995.
6. D. J. Campbell. Task complexity: A review and analysis. *The Academy of Management Review*, 13(1):40–52, 1988.
7. R. Capra, J. Arguello, A. Crescenzi, and E. Vardell. Differences in the use of search assistance for tasks of varying complexity. In *SIGIR*, pages 23–32. ACM, 2015.
8. B. J. Jansen, D. Booth, and B. Smith. Using the taxonomy of cognitive learning to model online searching. *IPM*, 45(6):643–663, 2009.
9. D. Kelly, J. Arguello, A. Edwards, and W.-c. Wu. Development and evaluation of search tasks for iir experiments using a cognitive complexity framework. In *ICTIR*, pages 101–110. ACM, 2015.
10. G. Keppel and T. D. Wickens. *Design and Analysis: A Researcher's Handbook.* Prentice Hall, 3 edition, 1991.
11. Y. Li and N. J. Belkin. A faceted approach to conceptualizing tasks in information seeking. *Information Processing and Management*, 44(6):1822 – 1837, 2008.
12. E. G. Toms. Task-based information searching and retrieval. In I. Ruthven and D. Kelly, editors, *Interactive information seeking, behaviour and retrieval*, chapter 3, pages 43–59. 2011.
13. P. Vakkari. Task complexity, problem structure and information actions. *Information Processing & Management*, 35(6):819–837, 1999.
14. P. Vakkari. Task-based information searching. *Annual Review of Information Science and Technology*, 37(1):413–464, 2003.
15. B. M. Wildemuth, L. Freund, and E. G. Toms. Untangling search task complexity and difficulty in the context of interactive information retrieval studies. *Journal of Documentation*, 70(6):1118–1140, 2014.