The Effects of Manipulating Task Determinability on Search Behaviors and Outcomes

Rob Capra
Jaime Arguello
Heather O'Brien
Yuan Li
Bogeum Choi

UNC SCHOOL OF INFORMATION AND LIBRARY SCIENCE

Search tasks

- ## Search tasks are a key part of interactive IR research
  - Used to evaluate systems and search behaviors
  - Understand task characteristics → user challenges
    (e.g., Byström & Hansen, 2005; Li & Belkin, 2008)
  - Inform the design of tools

- ## Task characteristics can influence:
  - Search behaviors and outcomes
  - User perceptions of search process

> **Our focus:**
> **Effects of task complexity**

Task complexity

- **Complexity is a multi-faceted concept:**
    - #paths/goals; uncertainty/interdependence (Campbell, 1988; Wildemuth et al., 2014)
    - cognitive complexity (Anderson & Krathwohl, 2001; Kelly et al., 2015)
    - *a priori* determinability/uncertainty (Byström & Järvelin, 1995; Bell & Ruthven, 2004)

- **Effects of task**
    - Complex tasks → greater difficulty/effort (Arguello 2014)
    - Degree of "challenge" (O'Brien)
        - Too easy → boredom; Too hard → disengagement

- **Our focus:**
    - Manipulate task scope to change task determinability
    - Measure the effects on search behaviors, outcomes, and perceptions

Task Manipulation

- **Comparative tasks:** consider *items* across *dimensions*

	objective dimensions		subjective dimensions	
Water purification	**Time required**	**Organisms eliminated**	**Taste**	**Safety vs. convenience**
Chemical tablets				**U**
Boiling				
Charcoal filters				

items

Unspecified (U): no items or dimension specified

- What are different methods for purifying water to drink from streams and how do they differ?

	objective dimensions		subjective dimensions	
Water purification	**Time required**	**Organisms eliminated**	**Taste**	**Safety vs. convenience**
Chemical tablets				U
Boiling				
Charcoal filters				

items

Task Manipulation

Items (I): specified two items to compare, but no dimension

- How do **boiling water** and using a **charcoal filter** differ as methods to purify stream water to drink?

	objective dimensions		subjective dimensions	
Water purification	**Time required**	**Organisms eliminated**	**Taste**	**Safety vs. convenience**
Chemical tablets				U
Boiling			I	
Charcoal filters				

items

Task Manipulation

Objective dimension (O): specified the dimension, but no items

- What are different methods for purifying water to drink from streams and how do they differ in terms of the **time required**?

Task Manipulation

Subjective dimension (S): specified the dimension, but no items

- What are different methods for purifying water to drink from streams and how do they differ in terms of the **water taste**?

Task Manipulation

Item + Objective Dimension (IO): specified items + an objective dimension

- How do **boiling** water and using a **charcoal filter** differ as methods to purify stream water to drink in terms of the **time required**?

Task Manipulation

Item + Subjective Dimension (IS): specified items + a subjective dimension

- How do **boiling** water and using a **charcoal filter** differ as methods to purify stream water to drink in terms of the **water taste**?

	objective dimensions		subjective dimensions	
Water purification	**Time required**	**Organisms eliminated**	**Taste**	**Safety vs. convenience**
Chemical tablets	O		S	U
Boiling	IO	I	IS	
Charcoal filters				

12 comparative tasks

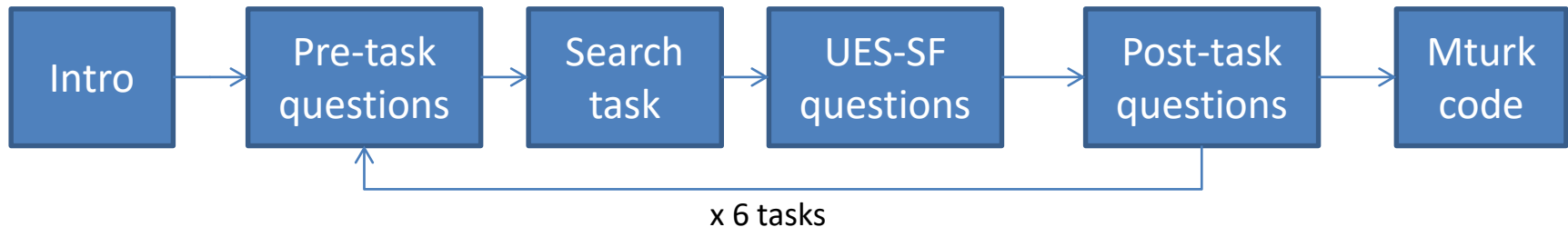**Table 1: Task topics, items, and dimensions used in our task descriptions.**

Topic	Items	Objective Dimension	Subjective Dimension
motor oil for cars	synthetic and organic oil	price range	cost-effectiveness
types of rice	white and brown rice	fiber content	noticeably affect insulin levels
types of ballet	classical and neo-classical	historical origin	difficulty of postures and movements
music speaker materials	polypropylene and paper	price	sound quality
garden fertilizers	organic and chemical	nutrient content	safety for growing vegetables
types of paint thinner	linseed and poppyseed oil	time for paint to dry	how well-suited for beginner
wifi routers	single band and dual band	signal interference	privacy and security issues
types of plastic	PET and PVC	how they can be recycled	risks involved in household use
indoor dog breeds	Pug and Bichon Frise	size at adulthood	ability to be left alone during the day
smoking cessation methods	nicotine gum and nicotine patch	average treatment length	difficulty to stop treatment
water purification methods	boiling and charcoal filter	micro-organisms eliminated	tradeoffs of safety and convenience
cooking skillet materials	aluminum and cast iron	how rapidly they heat up	what foods should/not be cooked

Research questions

Effects of task version on participants'…

- RQ1/RQ2: Task version → pre-/post-task perceptions?
    - Perceptions of determinability, subjectivity, knowledge, interest, difficulty.

- RQ3: Task version → level of engagement during the task?
    - User Engagement Scale short-form (O'Brien, et al., 2018)
        - 1) Focused attention, 2) aesthetic appeal, 3) perceived usability, 4) reward

- RQ4: Task version → search behaviors?
    - Measures of search effort, search diversity

- RQ5: Task version → search strategies?
    - Differences in querying strategies observed for different task versions

Method

- ## Within-subject user study (N=144)
  - Each participant completed 6 search tasks (U, I, O, S, IO, IS)
  - Custom search system (Bing API)

    U=unspecified, I=items,
    O=objdim, S=subjdim,
    IO=items+objdim, IS=items+subjdim

- ## Task goal
  - "Find and bookmark pages that would be useful for addressing the task"
  - + brief justification each bookmark

- ## Design
  - 12 task topics x 6 task versions = 72 combinations
  - Latin squares used to balance order of task version and task topics
  - 144 participants x 6 tasks = 864 search sessions

Study protocol

```
┌──────┐   ┌──────────┐   ┌────────┐   ┌──────────┐   ┌──────────┐   ┌────────┐
│Intro │ → │ Pre-task │ → │ Search │ → │  UES-SF  │ → │Post-task │ → │ Mturk  │
│      │   │questions │   │  task  │   │questions │   │questions │   │  code  │
└──────┘   └──────────┘   └────────┘   └──────────┘   └──────────┘   └────────┘
               ↑                                           │
               └───────────────────────────────────────────┘
                              x 6 tasks
```

- Crowdsourced user study

  - Amazon Mechanical Turk (Mturk)

  - U.S. workers with >= 95% acceptance rate

- Six tasks packaged as one Mturk HIT (USD $10)

- 70-90 minutes total

- System allowed participants to pause/resume between tasks

- Questionnaires used the similar statements pre/post
  1 = strongly disagree … 5 = strongly agree

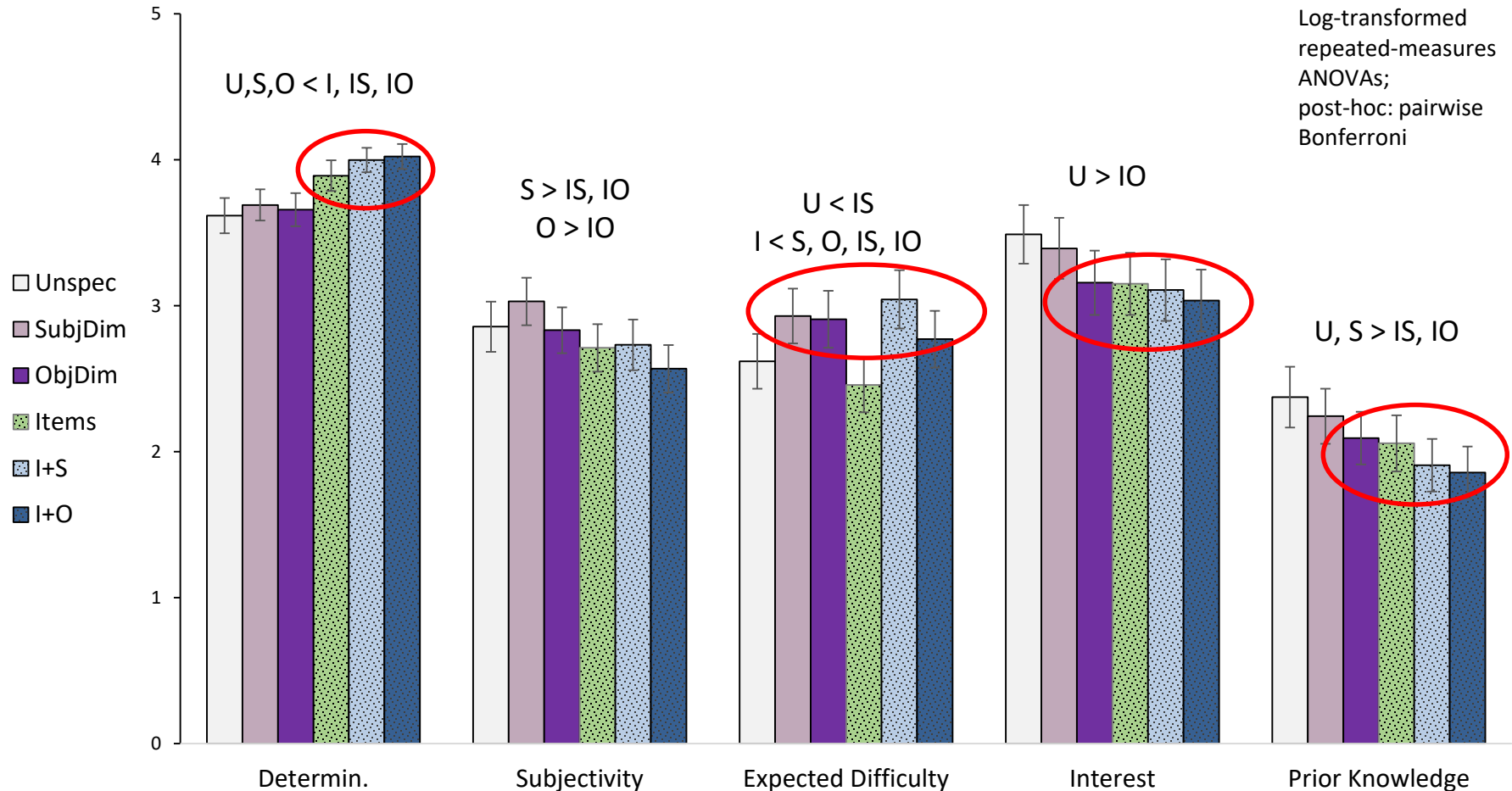| **Knowledge**	• I already know a lot about this topic.
Interest	• I am interested in the topic.
Difficulty	• I think the task will be difficult.
Determinability	• The task description tells me exact items that I need to compare. • The task description tells me exact criteria that I need to consider in understanding the differences between items. • The task description has details that will help me complete the task. • The task is specific. • The information requested in narrowly focused. • Right now, I know some specific things to look for to address the task. • The task description provides me with new information that I did not already know about this topic.
Subjectivity	• The task will require gathering information regarding people's feelings, tastes, and/or opinions. • The task is open-ended.

RQ1: Pre-task perceptions

- Specifying items (I, IS, IO) → increased determinability
- Specifying dimensions (S, O, IS, IO) → increased expected difficulty
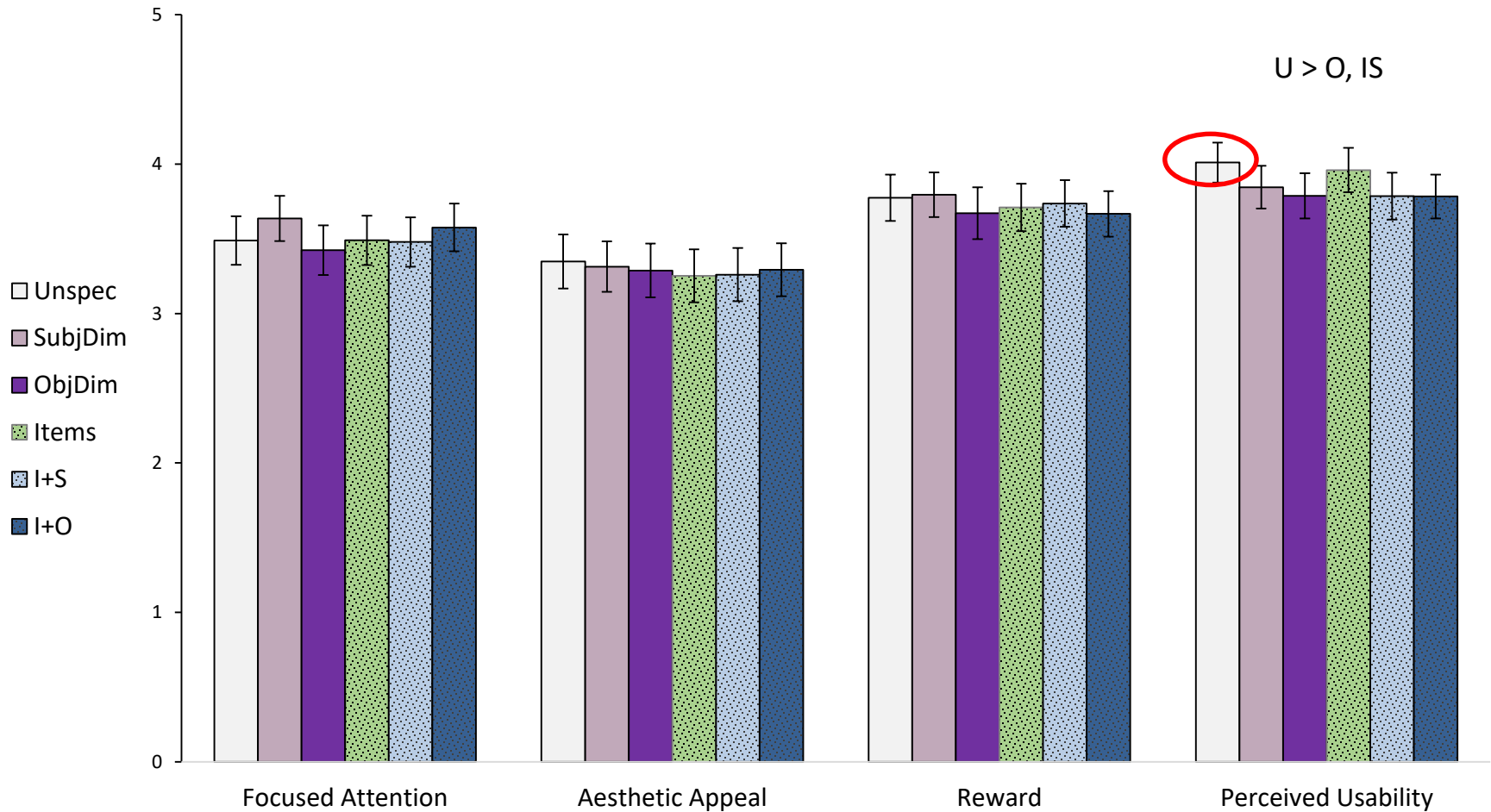- Items & dims → lower interest & prior knowledge

Log-transformed repeated-measures ANOVAs;
post-hoc: pairwise Bonferroni

U,S,O < I, IS, IO

S > IS, IO
O > IO

U < IS
I < S, O, IS, IO

U > IO

U, S > IS, IO

Legend:
- Unspec
- SubjDim
- ObjDim
- Items
- I+S
- I+O

X-axis categories: Determin. | Subjectivity | Expected Difficulty | Interest | Prior Knowledge

- Specifying items (I, IS, IO) → increased determinability (same as pre)
- Specifying dimensions (S, O, IS, IO) → increased difficulty (same as pre)
- No observed effect on interest increase & knowledge increase
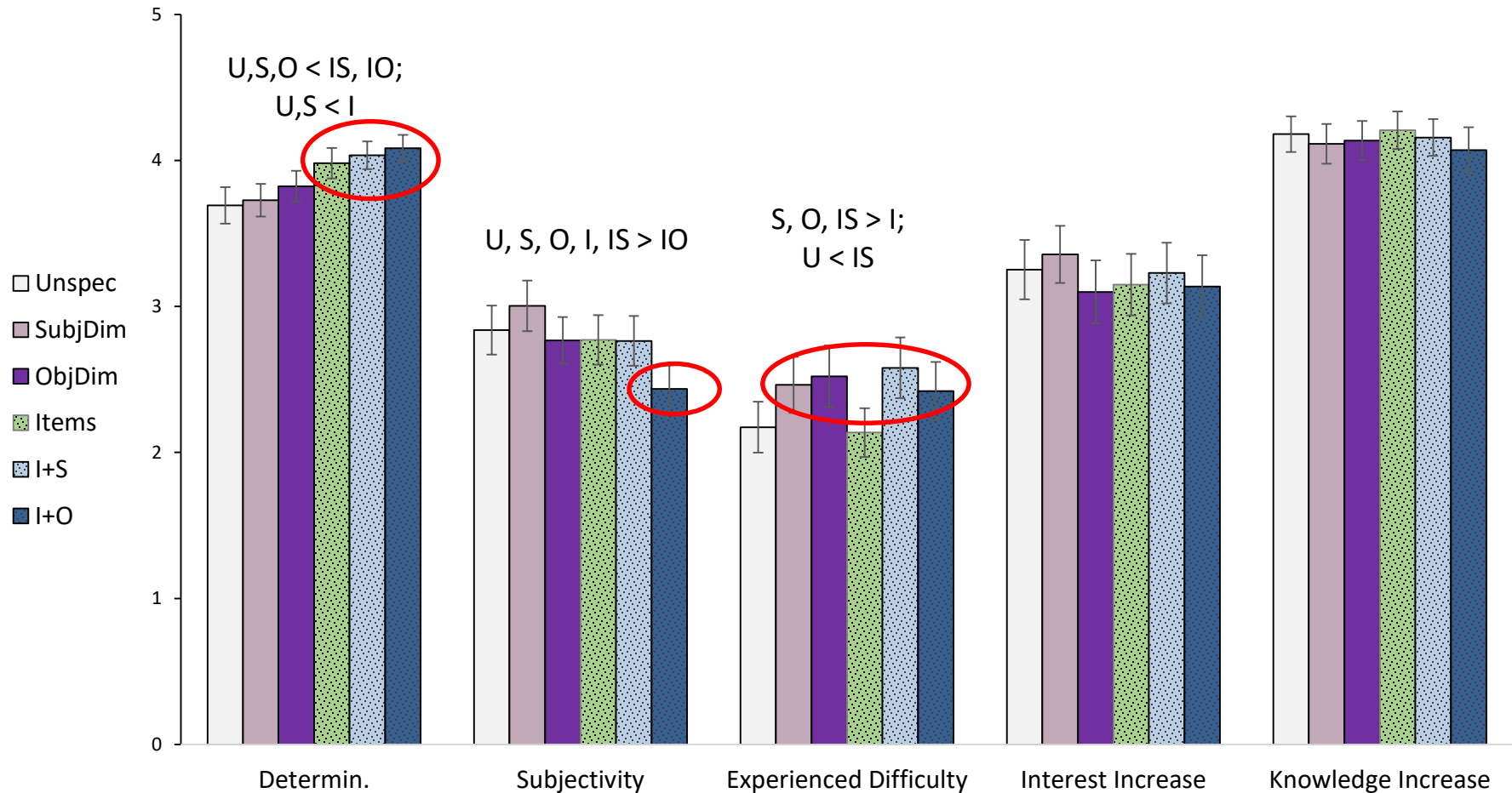
RQ3: Post-task engagement

- No observed effect on focused attention, aesthetic appeal, or reward
- Unspecified had higher perceived usability than O or IS

- Unspecified
  - least effort, similar search strategies
  - → satisficing?

- Items only
  - low effort, few unique behaviors
  - items are concrete concepts (nouns)
  - consistent terms in queries and relevant docs

- Dimensions (O or S)
  - greater effort, divergent strategies
  - dims are more abstract (even O)
  - more difficult to generate query terms
  - more difficulty to identify in docs
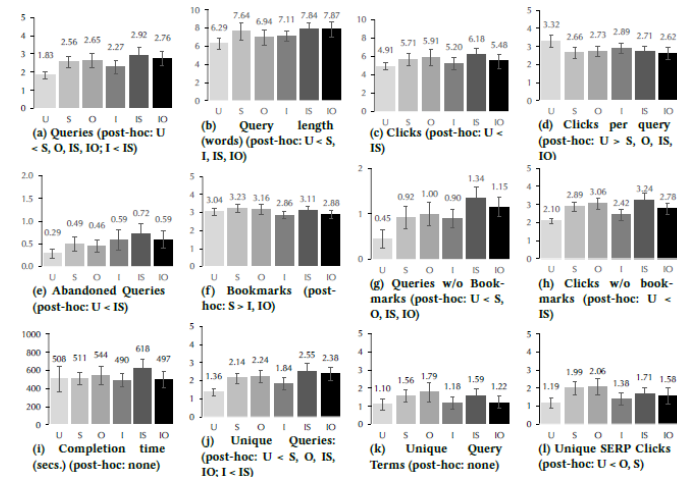
Measures:
- #queries
- query length
- #clicks
- clicks/query
- abandoned queries
- #bookmarks
- queries w/o book
- clicks w/o book
- completion time
- #uniq queries
- uniq query terms
- uniq SERP clicks

Table 4: Example codes based on items and dims.

Query	Item(s)	Dim(s)
methods to quit smoking	*	*
how do smoking cessation methods differ	*	*
nicotine gum vs patches	I	*
average treatment length in quitting smoking	*	D
average nicotine patch *treatment length*	I	D

RQ5: Search Strategies

	(*,*)	(I,*)	(*,D)	(I,D)
U	0.937 (0.192)	0.020 (0.092)	0.044 (0.171)	0.000 (0.000)
S	0.413 (0.368)	0.004 (0.026)	0.558 (0.371)	0.025 (0.111)
O	0.407 (0.365)	0.012 (0.061)	0.567 (0.368)	0.013 (0.059)
I	0.037 (0.114)	0.915 (0.211)	0.002 (0.023)	0.046 (0.179)
IS	0.018 (0.081)	0.417 (0.406)	0.051 (0.176)	0.513 (0.420)
IO	0.016 (0.101)	0.318 (0.396)	0.035 (0.132)	0.631 (0.403)

- Unspecified (U) → 94% did not include items or dims
- Dims (S and O) → split between *,* and *,D
- Items (I) → 91% were queries for items only
- IS and IO → split between I,* and I,D

Summary:

- **items are easy to search for**
- **dimensions add complexity, are more difficult**

Summary

- **Unspecified**
  - Lowest determinability, but easiest, least diversity
  - → satisficing or self-defining a scope to reduce uncertainty

- **Items**
  - Items are concrete concepts
  - Narrowed scope, reduced #outcomes, reduced uncertainty
  - → easy to query and identify in relevant docs (low #uniq queries, low #query terms)

- **Dimensions**
  - Dimensions (both O and S) are more abstract concepts (vocabulary gap)
  - Narrowed scope, but more challenging/difficult
  - → more search effort, more diverse search behaviors

- **Objective vs. Subjective Dimensions**
  - Did not observe strong differences
  - Trend: subjective increased search effort

Implications

…for IIR studies

- Complex relationship: scope, determinability, difficulty
    - Broad scope, low determinability → could be easy task
    - Narrow scope, high determinability → could be difficult task
- Items and dimensions
    - Items can make a task less complex → decrease uncertainty
    - Dimensions can make a task more complex → increase uncertainty
    - Consider both **outcomes** and **processes**

…for tools

- Dimensions created complexity, so…
    - Algorithmically infer dimensions of current task
    - Incorporate dimensions into query suggestions, faceted search
    - Grid/matrix interface to help users with comparative tasks

Questions?

Rob Capra, Jaime Arguello, Heather O'Brien, Yuan Li, Bogeum Choi

Email: rcapra3@unc.edu
Twitter: @rcapra3

Backup slides

User Engagement Scale (short-form)

- Based on O'Brien's 31-item User Engagement Scale
- Short-form has 12 questions along four dimensions

Focused Attention	• I lost myself in this search experience. • The time I spent searching just slipped away. • I was absorbed in the search task.
Perceived Usability	• I felt frustrated while doing the search task. • My search experience was taxing. • I found the search system confusing to use.
Aesthetic Appeal	• The search system was attractive. • The search system was aesthetically appealing. • The search system appealed to my senses.
Reward	• My search experience was worthwhile. • My search experience was rewarding. • I felt interested in the search task.