

The Effects of Manipulating Task Determinability on Search Behaviors and Outcomes

Robert Capra¹, Jaime Arguello¹, Heather O'Brien², Yuan Li¹, Bogeum Choi¹

¹University of North Carolina at Chapel Hill

²University of British Columbia

{rcapra,jarguello}@unc.edu,h.obrien@ubc.ca,{yuanli,choiboge}@live.unc.edu

SIGIR '18: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval



2018 Proceeding

General Chairs:  [Kevyn Collins-Thompson](#),  [Qiaozhu Mei](#), [+ 3](#)

Publisher: Association for Computing Machinery, New York, NY, United States

Conference: SIGIR '18: The 41st International ACM SIGIR conference on research and development in Information Retrieval • Ann Arbor MI USA • July 8 – 12, 2018

ISBN: 978-1-4503-5657-2

Published: 27 June 2018

Sponsors: [SIGIR](#)

 [Recommend ACM DL](#)

ALREADY A SUBSCRIBER? [SIGN IN](#)

 [Get Alerts for this Conference](#)

 [Save to Binder](#)

 [Export Citation](#)

アジェンダ

1. 導入
2. 実験手法
3. 結果
4. ディスカッション

1.導入

導入：目的

検索タスクにおける決定可能性の操作の影響を確かめること

特に以下の項目を論文中で確認

- 決定可能性の印象
- エンゲージメント
- 検索時の行動
- 検索戦略

導入：検索タスクの重要性

検索タスク (Search Task^{*1}) は、IIR^{*2}研究の中心

なぜか？

1. 研究の手段として役に立つ
- 2. 研究の対象としても扱われ、
ユーザーをサポートする新たなツールの設計に役に立つ**

^{*1} 検索タスク(Search Task):

ユーザーが特定の目標を達成するために情報を探す活動

タスクの構造、複雑さ、決定可能性が異なり、ユーザーがどのようにアプローチしタスクを完了するかに影響する

^{*2} IIR(Interactive Information Retrieval):

ユーザーが特定の検索タスクを達成するために検索システムとやりとりする枠組み

導入：検索タスク次元

検索タスクは二つの次元によって特徴付けられる

1. 一般的な側面

- 自発的 or 強制的

2. 共通属性

- 客観的な属性
 - タスクの複雑さ etc.
- 主観的な属性
 - ユーザーの専門知識 etc.

導入：検索タスク研究の焦点

多くの研究が、

検索タスクをどのように変更できるか、に焦点が当ててきた

Ex.

- 検索タスクの主な活動の変更
 - 検索 vs ブラウジング
- 検索タスクの目標
 - 明確な目標 vs 曖昧な目標
- タスクの構造
 - **タスクの複雑さ**

導入：検索タスクの複雑さ

- 検索タスクの**複雑さ**_{*1}は近年注目されており、**検索行動や検索結果に影響を与える**ことがわかっている

*1 複雑さ(Task Complexity):

必要な結果の数やそれに至る経路の数、不確実性の程度など、さまざま要因によって決まる
複雑なタスクには複数の解決策が存在する

導入 : 複雑さと決定可能性

Byström and Järvelin (1995)はタスクの複雑さを決定可能性^{*1}という概念で測定することを提案

彼らは決定可能性を

「**検索者がタスクを内在化し、以下を推測できる程度**」と定義

1. 必要な結果
2. 結果を生成するために必要な情報
3. 必要な情報を集めるための手順

^{*1} 決定可能性(Determinability)

タスク完了までに必要な結果やプロセスに関する不確実性の度合い

導入：決定可能性

Wildemuth et al. (2014)の文献研究を基に、タスクの複雑性を以下三つに定義

1. ステップの多様性

- 東京タワーの所在地 vs 東京タワーの作られた経緯、建設過程、技術を調査

2. 概念の多様性

- 健康に良い運動法 vs 運動法を医療的、心理的、社会的視点から調査

3. 決定可能性

- 不確実性の程度

導入：検索行動と結果

タスクの複雑さが、影響を与える検索行動・検索結果

- タスクを行う前のタスクの難易度評定 (Bell & Ruthven, 2004; Capra, Arguello, Crescenzi, & Vardell, 2015; Hu & Kando, 2017; Kelly et al., 2009; Wu, Kelly, Edwards, & Arguello, 2012)
- 実際の難易度評定 (Byström & Järvelin, 1995; Bell & Ruthven, 2004; Capra, Arguello, Crescenzi, & Vardell, 2015; Hu & Kando, 2017; Kelly et al., 2009; Wu, Kelly, Edwards, & Arguello, 2012)
- 検索努力^{*1} (Byström & Järvelin, 1995; Capra, Arguello, Crescenzi, & Vardell, 2015; Hu & Kando, 2017; Jansen, Booth, & Smith, 2009; Kelly et al., 2009; Wu, Kelly, Edwards, & Arguello, 2012)
- さまざまなクエリ^{*2}の再構築 (Wildemuth, Kelly, Boettcher, Moore, & Dimitrova, 2018)

*1 検索努力:

ユーザーが情報を見つけるためにクエリの入力、結果のクリック、結果ページの確認等に費やす行動や労力

*2 クエリ(Query):

検索タスクを実行する際にユーザーが入力する検索キーワードやフレーズのこと

導入 : エンゲージメント-1

タスクの特性は、
ユーザーのエンゲージメント^{*1}にも影響を与える可能性がある

エンゲージメント研究は、以下を対象にしたものが多い

- システム特性
- 情報コンテンツ

↓

タスク特性のユーザーエンゲージメントへの影響に関する研究は希薄

^{*1} エンゲージメント：
検索を行う際のユーザーの認知的・感情的・時間的な投資の量(O'Brien, H. L., 2016)

導入 : エンゲージメント-2

タスクの難易度がユーザーエンゲージメントの重要な構成要素の一つ (O'Brien H. L., and Toms E. G., 2008)

Kelly et al. (2013) では、参加者が複雑さの高いタスクに対して、高いエンゲージメントを報告

↓

タスク操作が、

参加者がタスク完了時に感じる難易度に影響を与え、
エンゲージメントにも影響を与える可能性がある

導入：目的

検索タスクにおける決定可能性の操作の影響を確かめること

↓

- 考慮すべき特定のアイテム/次元を追加/除外し、
タスクの範囲を広げたり狭めたりすることによって間接的に操作
 - タスク特性を操作するために、比較検索タスク_{*1}を採用
- エンゲージメントへの影響も確認

*1 比較検索タスク(Comparative Search Task):
ユーザーが一連の次元の一連のアイテムを比較するタスク

導入：リサーチクエスション

RQ1. タスクの種類に参加者の**プレアンケートへの影響**はどのようなものか

RQ2. タスクの種類に参加者の**ポストアンケートへの影響**はどのようなものか

RQ3. タスクの種類に参加者の**タスク中のエンゲージメントへの影響**はどのようなものか

RQ4. タスクの種類**の検索行動への影響**はどのようなものか

RQ5. タスクの種類**の検索戦略への影響**はどのようなものか

2.実験手法

実験手法：実験内容

実験内容：

- 特注の検索システムを用いて、**指定された情報を探すのに便利なWebページをブックマークすること**
 - 検索システムは、Bing Web Search APIの検索結果を表示した
- **ブックマークした理由を説明すること**

実験手法：被験者

被験者：

- 144人
- 被験者間実験
- MTurkを利用
- アメリカ在住者を対象
- 70-90mの実験参加→\$10
- 参加許可率95%未満
- 実験の参加には、
Webブラウザに**Bookmarklet**の追加が求められた

実験手法：Bookmarklet

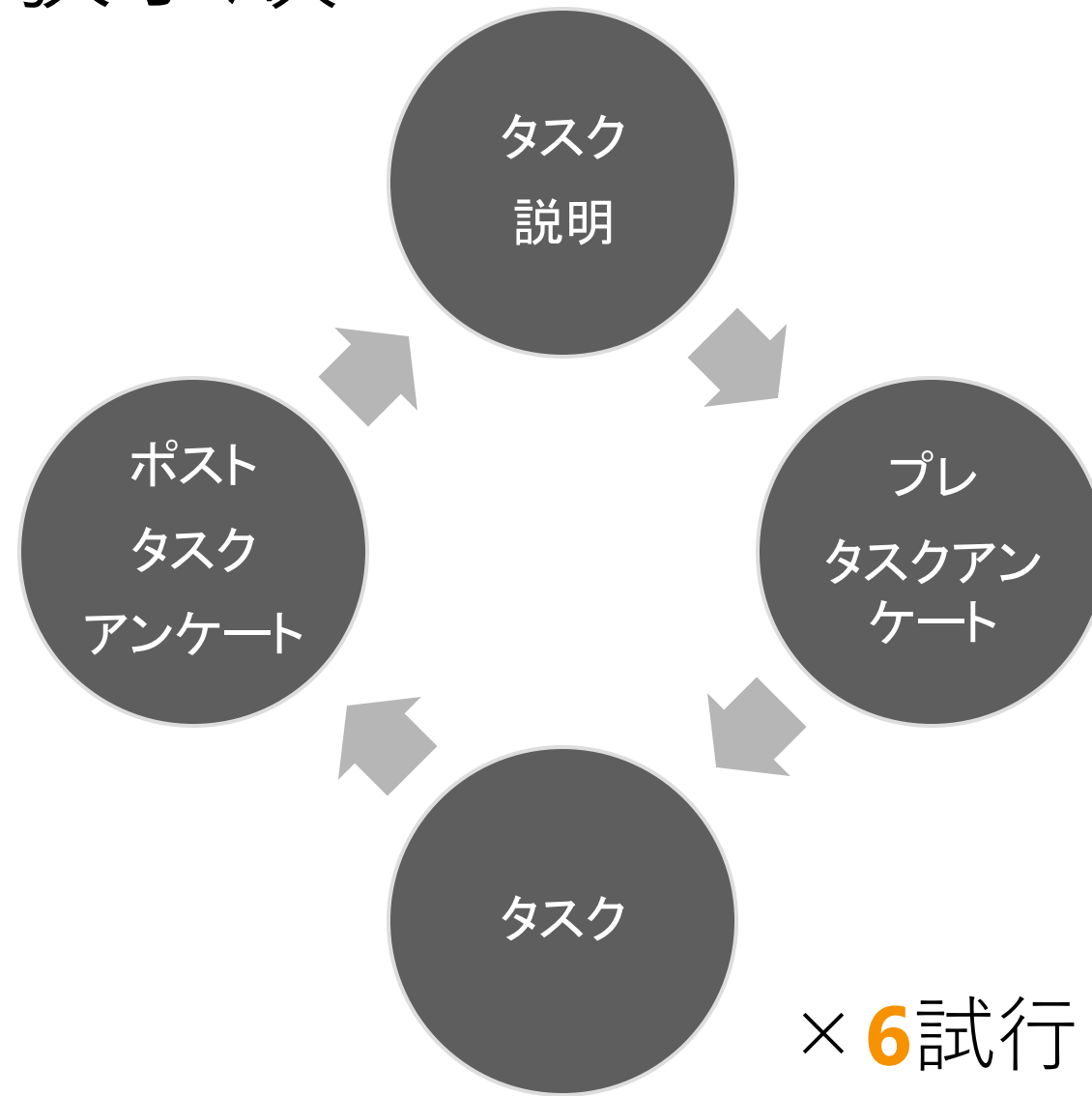
Bookmarkletを用いてできること：

1. ページのブックマークとその理由の記載
2. ブックマークしたページリストの閲覧とリストからの削除
3. SERP^{*1}に戻ることに
4. タスク完了時にタスク完了を知らせること

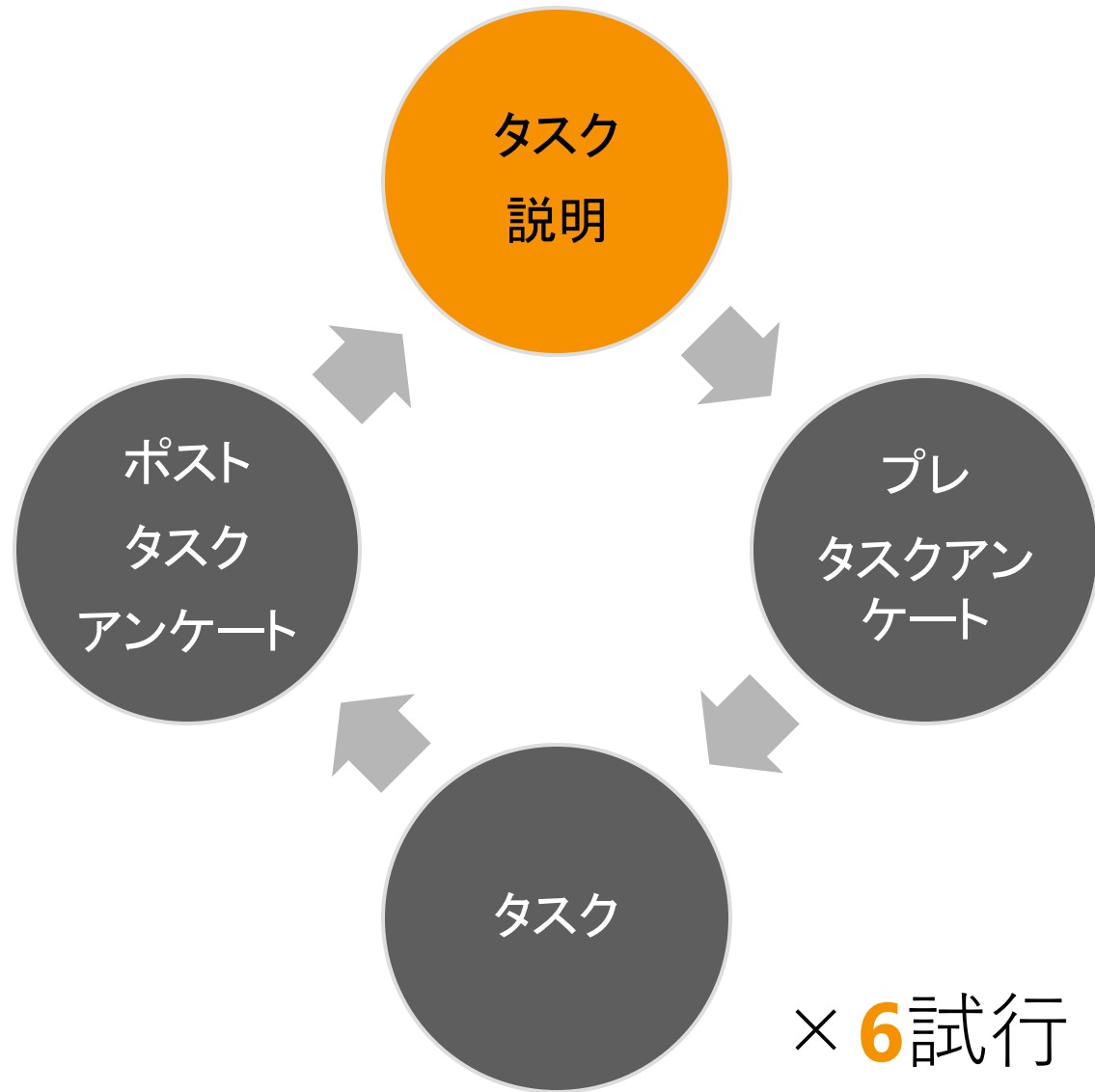
*1 SERP(Search Engine Result Page):

ユーザーが検索エンジンで検索した結果が表示されるページ

実験手法：実験手順



実験手法：タスク説明



タスク説明：

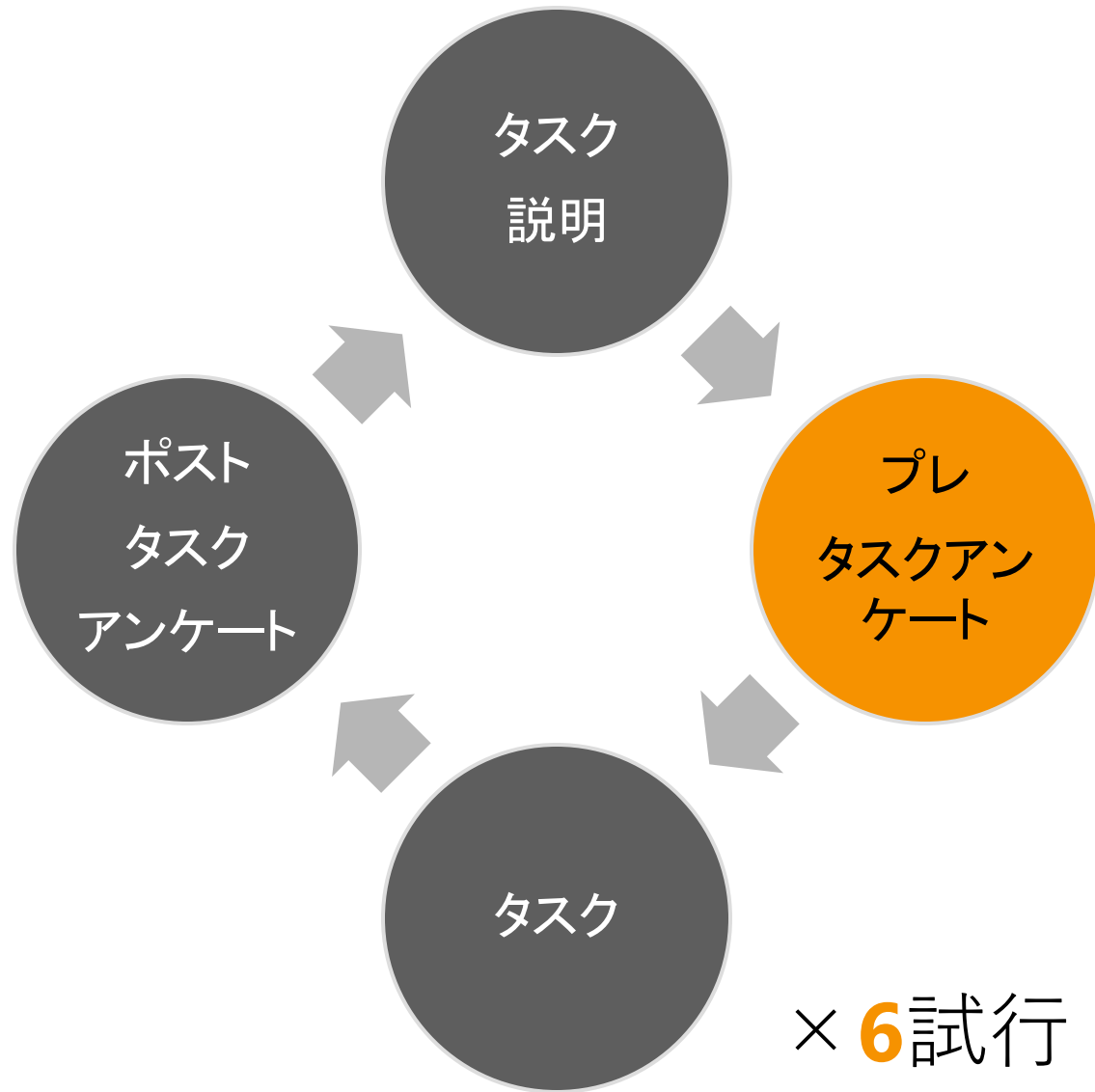
- 与えられた課題に回答するのに必要な **Webページのブックマーク** を目的としていること
- **ブックマークした理由** を説明する必要があること
- 必要なページを必要なだけブックマークすること

を毎タスク前に説明

実験手法：プレタスクアンケート-1

プレタスクアンケート：

- 14の質問
 - 5件法
 - ポストアンケートでも類似の内容を尋ねた



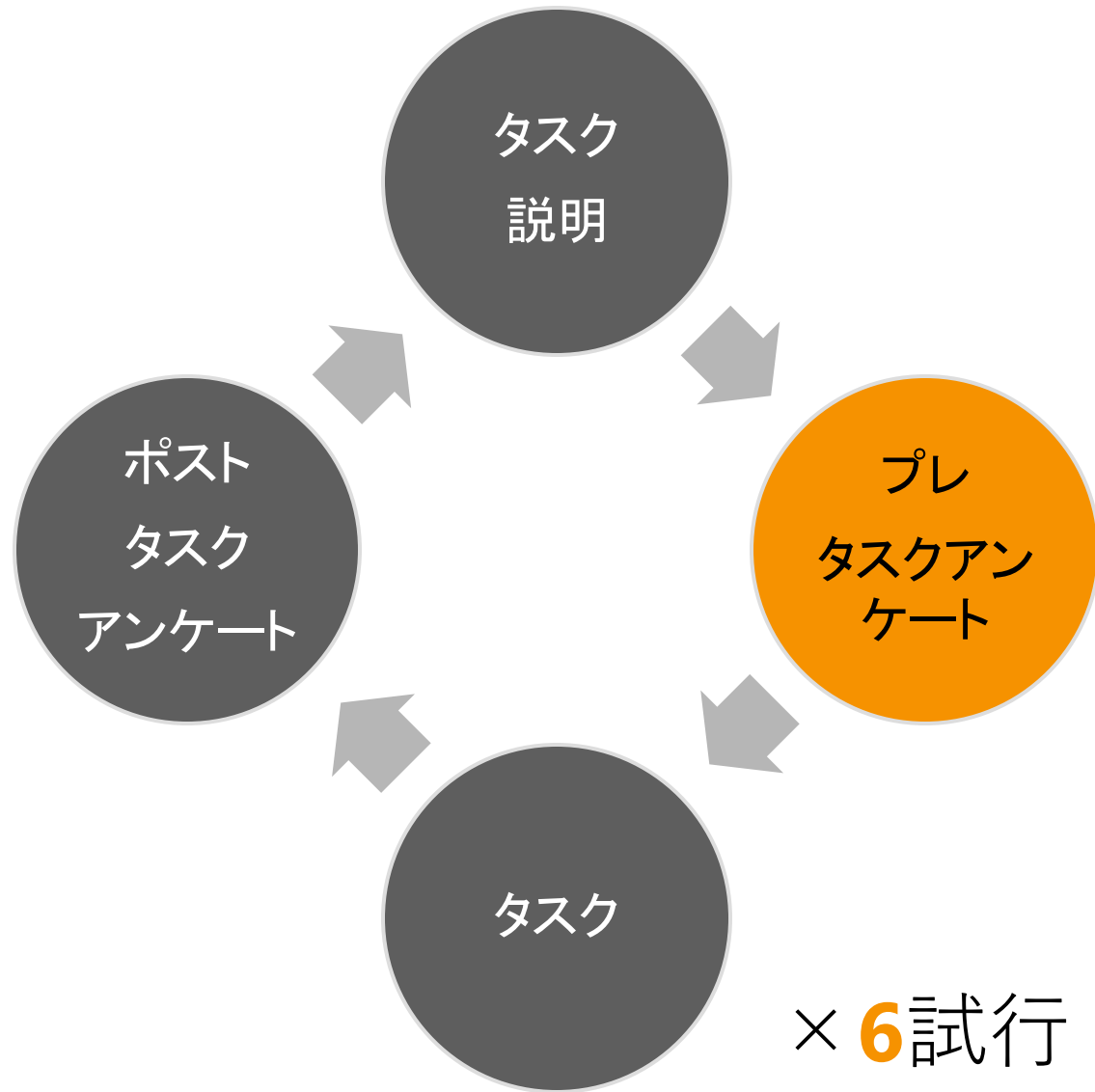
実験手法：プレタスクアンケート-2

プレタスクアンケート：

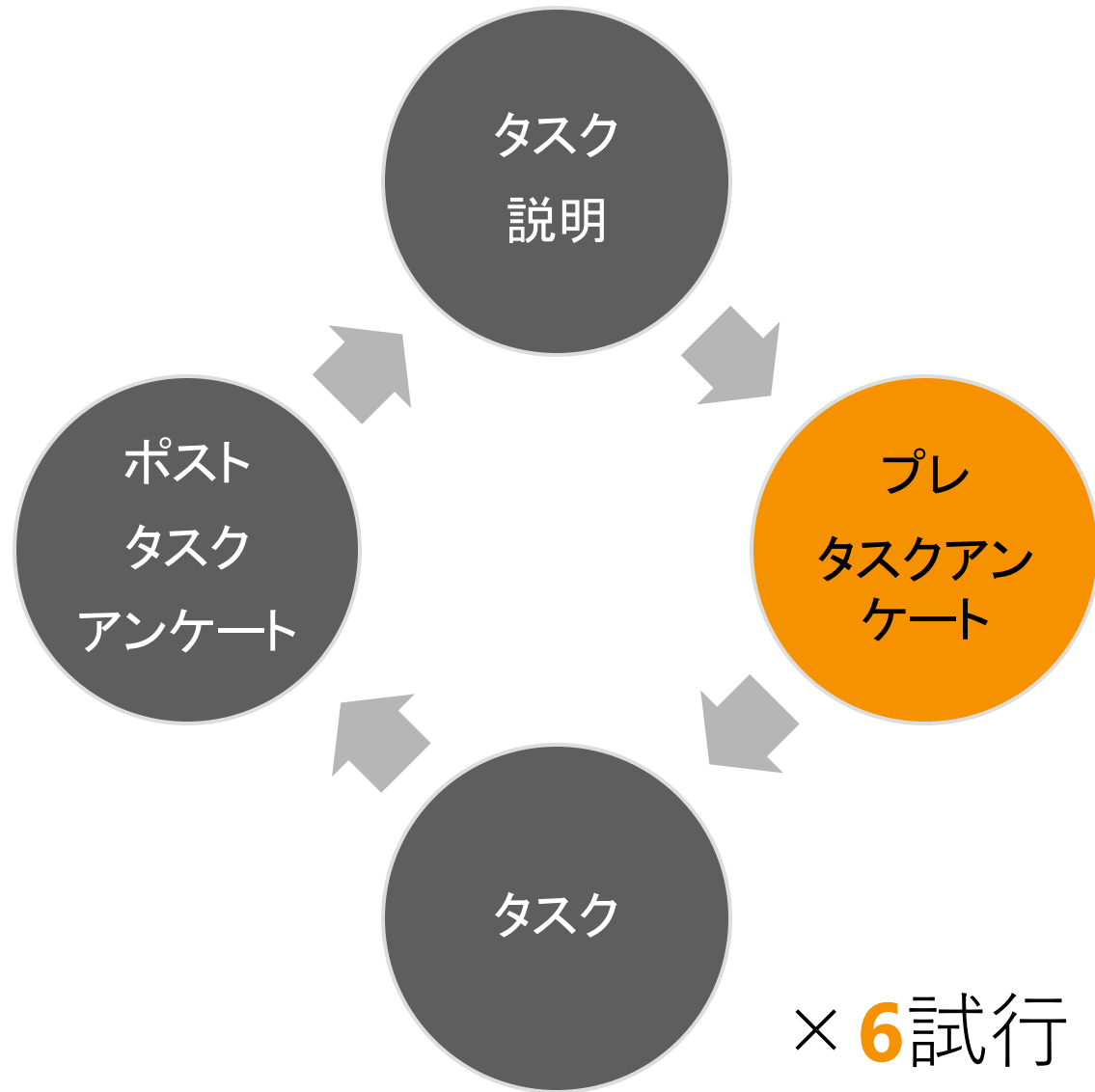
- 質問は5つの測定項目で構成
 1. 決定可能性 (Determinability)
 2. 主観性 (Subjectivity)
 3. 予想難易度 (Expected Difficulty)
 4. 事前興味 (Interest)
 5. 事前知識 (Prior Knowledge)

} 新規内容
計**11**項目

} 各**1**問



実験手法: プレタスクアンケート-3



決定可能性, 主観性:

- 11項目に対して、オブリミン法因子分析を実施
↓
 - D7項目を含む1因子を特定
 - Chronbach' α 事前=0.77, 事後=0.81
と許容範囲
 - S2項目を含む1因子を特定
 - Chronbach' α 事前=0.48, 事後=0.54
と低め
- 各項目の因子負荷量が0.45と十分なため使用

D:決定可能性 / S:主観性

実験手法：プレタスクアンケート-4

事前知識
事前興味
予想難易度

決定可能性

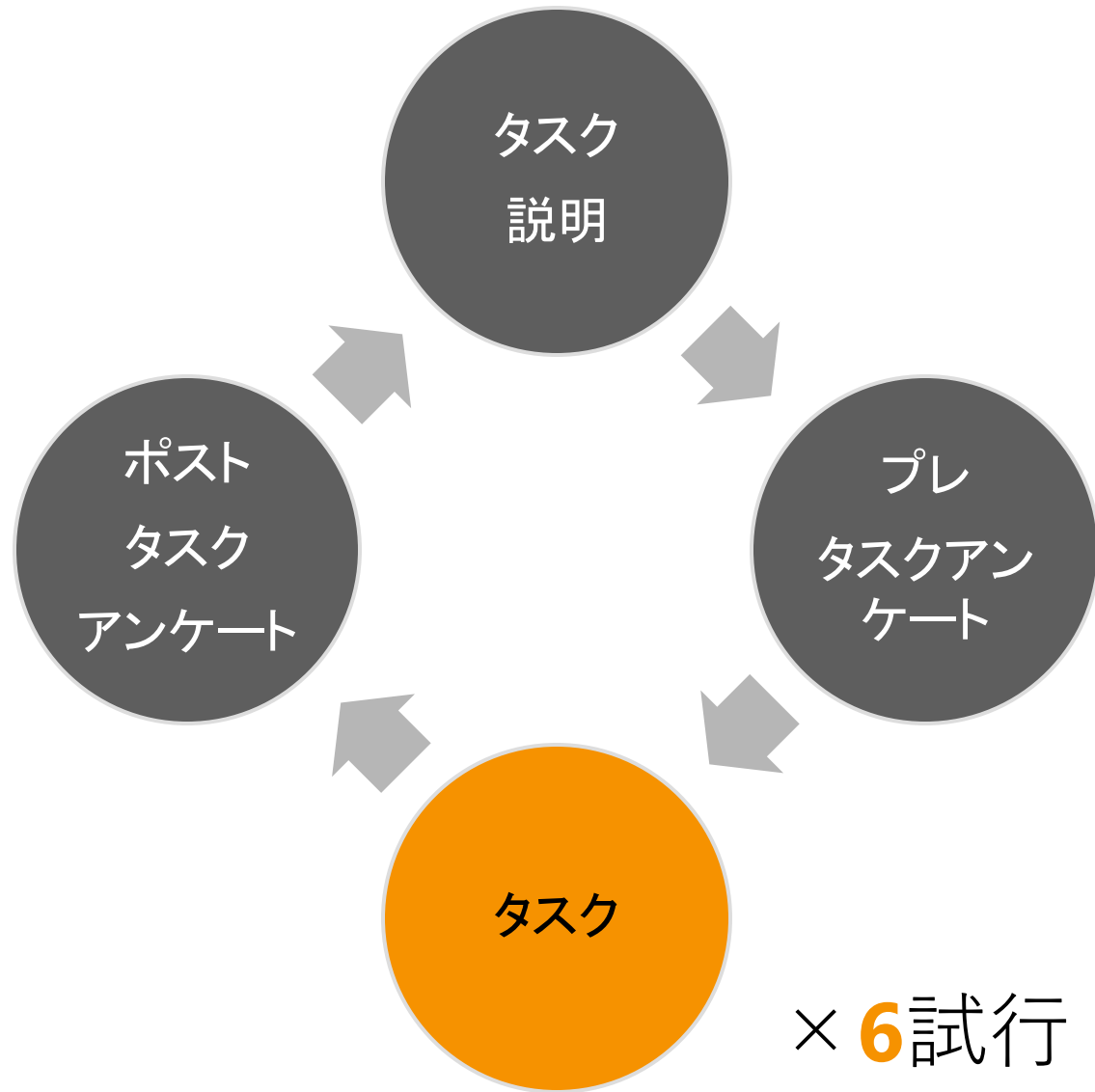
主観性

Measure	Pre-task question
knowledge	I already know a lot about this topic.
interest	I am interested in the topic.
difficulty	I think the task will be difficult.
specificity	The task is specific.
narrow	The information requested is narrowly focused.
newinfo	The task description provides me with new information that I did not already know about this topic.
figureout	I will need to figure out things that are not specified in the task description.
details	The task description has details that will help me complete the task.
lookfor	Right now, I know some specific things to look for to address the task.
exactitems	The task description tells me exact items that I need to compare.
exactdims	The task description tells me exact criteria that I need to consider in understanding the differences between items.
openended	The task is open-ended.
opinions	The task will require gathering information regarding people's feelings, tastes, and/or opinions.
factualinfo	The task will require gathering factual information.

Table2

プレで使用されたアンケート項目

実験手法：タスク-1



タスク：

- 特注の検索システムを用いて、**指定された情報を探すのに便利なWebページをブックマークすること**
- **ブックマークした理由を説明すること**
- 12トピック×6種類のタスクで構成

実験手法：タスク-2

12タスクトピック × 6種類のタスク

トピック	アイテム	客観的次元	主観的次元
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

Table1

実験にて使用されたタスクトピック、アイテム、ディメンション

実験手法：タスク-3

12タスクトピック × **6種類のタスク**

- アイテムの提示の有無
- 次元の限定の有無
 - 客観的次元
 - **事実**に基づいた情報
 - 主観的次元
 - **感情や知覚、意見、経験**に関する情報

これらの操作により
6種のタスクを用意

実験手法：タスク-5

12タスクトピック × **6種類のタスク**

条件	アイテムの限定の有無	次元の限定の有無と種類
Unspecified : U	×	×
Items : I	○	×
Objective dimension : O	×	Obj
Subjective dimension : S	×	Sub
Items + Objective dimension. IO	○	Obj
Items + Subjective dimension : IS	○	Sub

実験手法：タスク-4

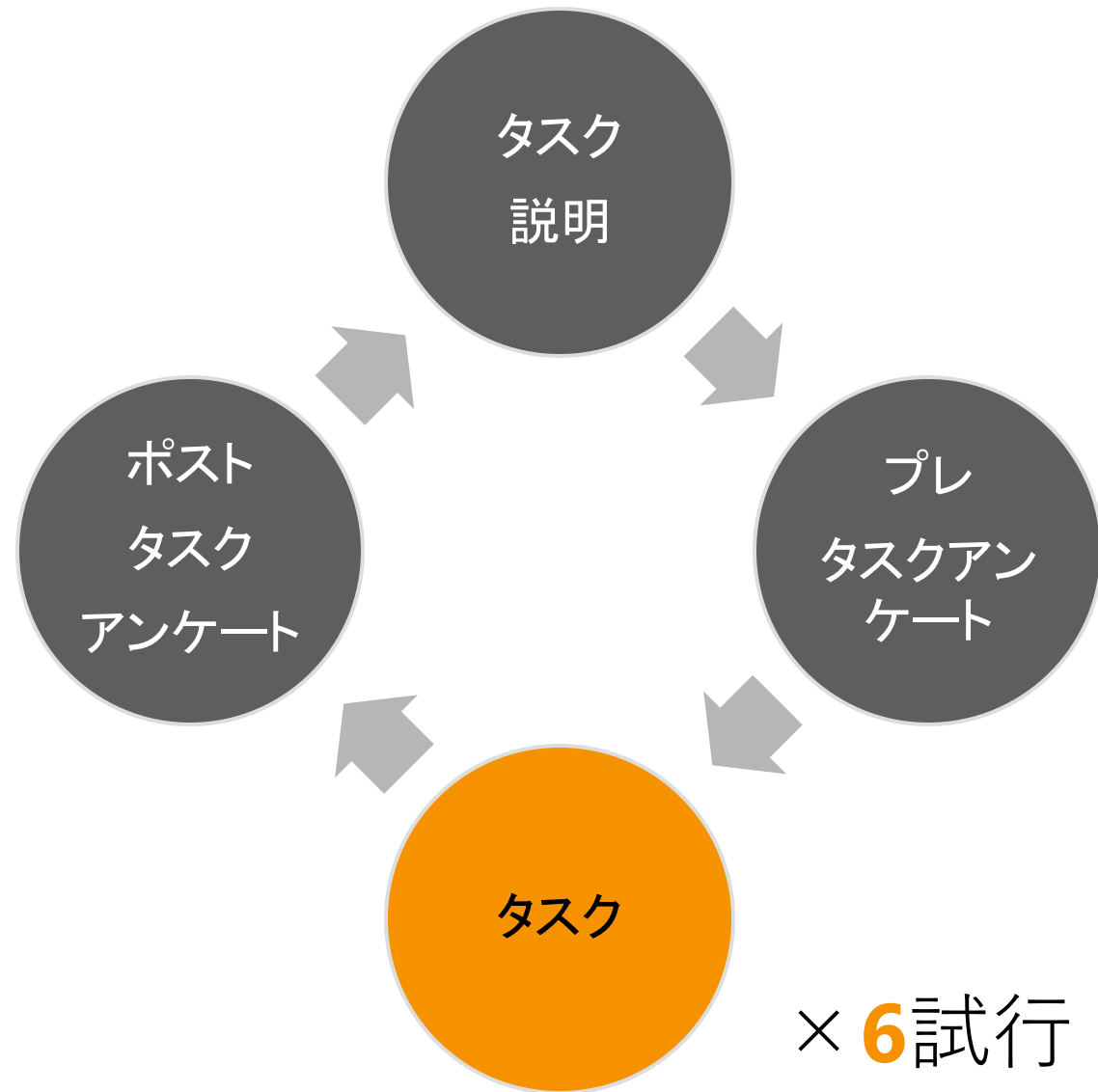
12タスクトピック × 6種類のタスク

トピック	アイテム	客観的次元	主観的次元
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

Table1

実験にて使用されたタスクトピック、アイテム、ディメンション

実験手法：タスク-6



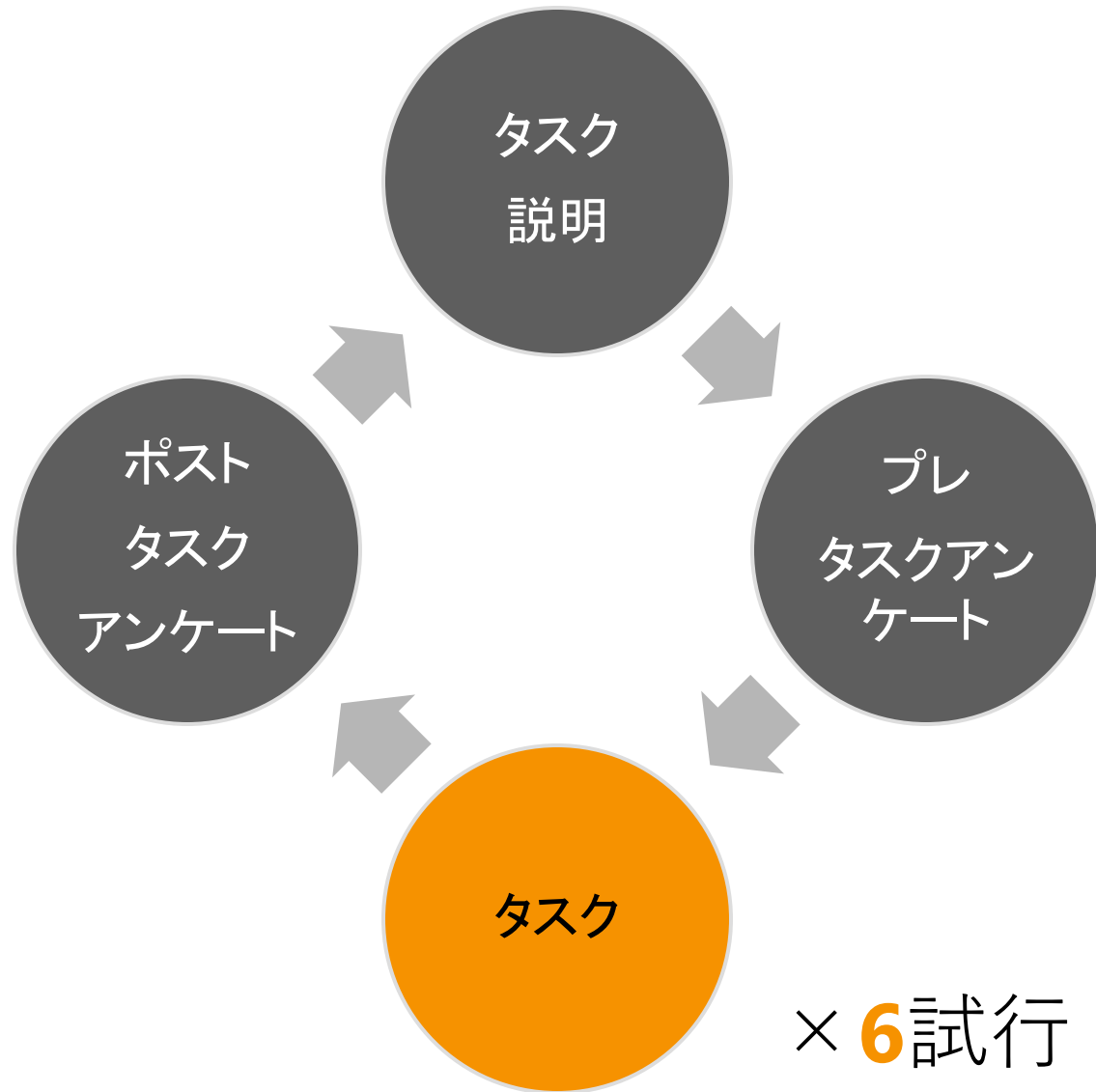
検索行動に関する測定項目：

1. クエリ*1数
(number of queries)
2. クエリの平均の長さ（単語の数）
(average query length (in words))
3. クリック数(number of clicks)
4. 1クエリあたりのクリック数
(number of clicks per query)
5. 何もクリックせず次のクエリに進んだクエリ数
(number of abandoned queries)
6. ブックマークされたサイトの数
(number of bookmarks)
7. ブックマークされていないクエリ数
(number of queries without a bookmark)
8. ブックマークをしていないクリック数
(number of clicks without a bookmark)
9. タスク完了までにかかった時間
(task completion time)

*1 クエリ(Query):

検索タスクを実行する際にユーザーが入力する検索キーワードやフレーズのこと

実験手法：タスク-7



検索行動に関する測定項目：

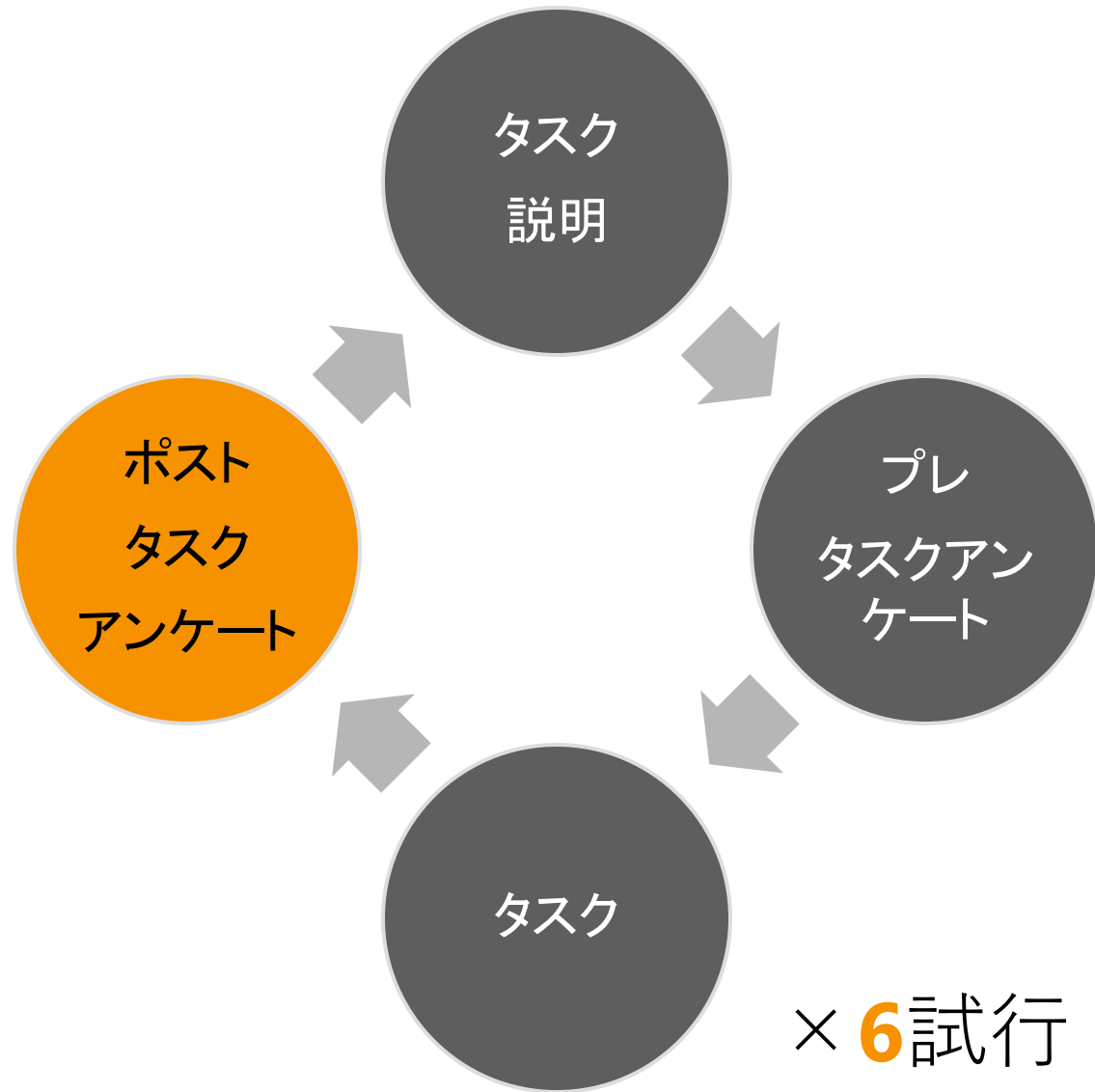
参加者が
他の同条件の参加者とどの程度違うか
を確認するために以下を追加で使用

1. ユニークなクエリの数
(number of unique queries
= not issued by any other participant)
2. ユニークなクエリ語の数
(number of unique query terms
(not used by any other participant))
3. SERP*₁上でクリックされたユニークなURLの数
(number of unique URLs clicked on a SERP
= not clicked by any other participant)

*1 SERP(Search Engine Result Page):

ユーザーが検索エンジンで検索した結果が表示されるページ

実験手法：ポストタスクアンケート-1



ポストタスクアンケート：

- 質問は5つの測定項目で構成
 1. 決定可能性
 2. 主観性
 3. 実際の難易度
 4. 興味の増加
 5. 知識の増加

+

簡易版

ユーザー・エンゲージメント尺度

実験手法：ポストタスクアンケート-2

知識増加
興味増加
実際の難易度

決定可能性

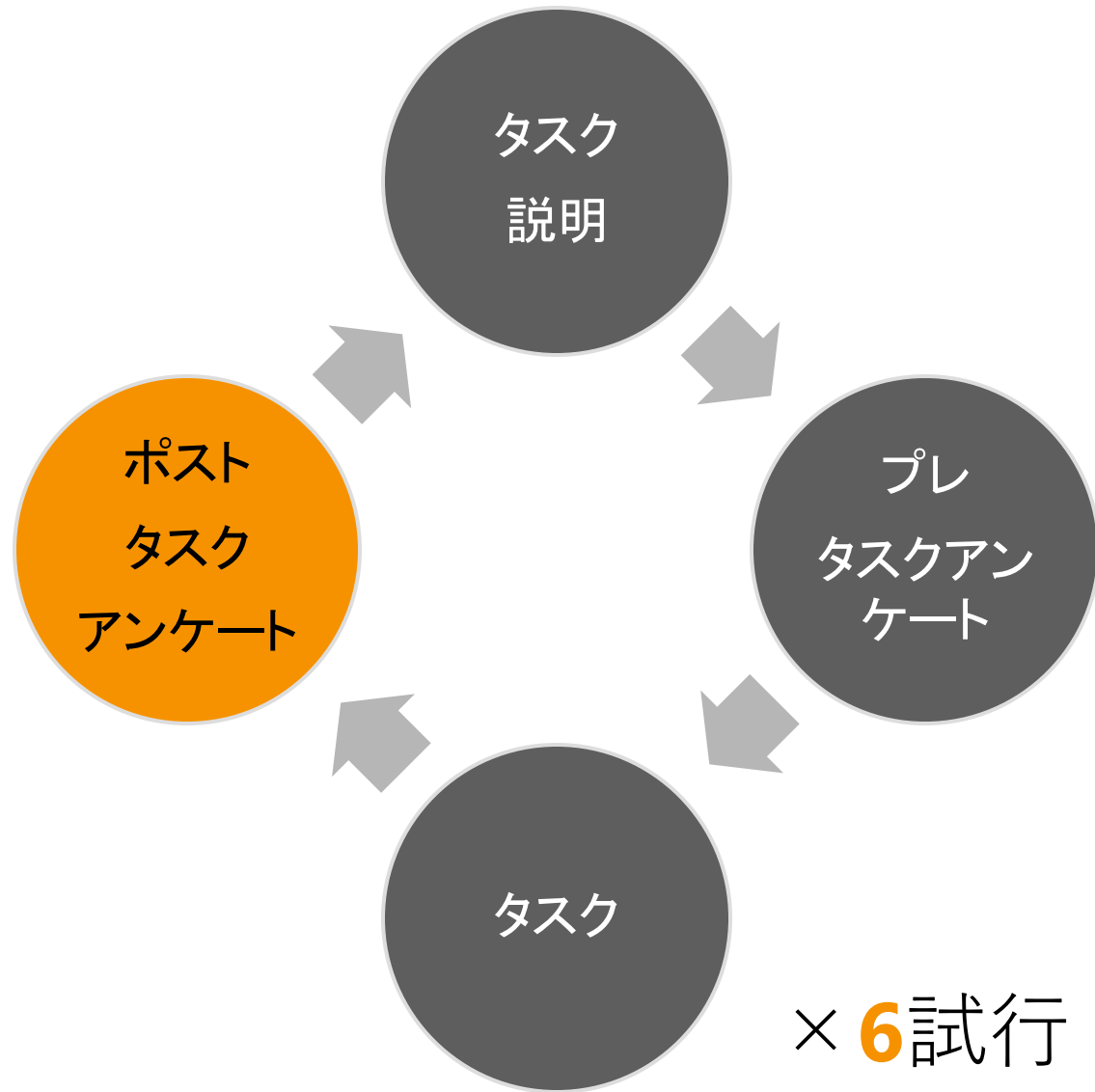
主観性

Measure	Post-task question
knowledge	My knowledge of this topic has increased.
interest	My interest in the topic has increased.
difficulty	The task was difficult.
specificity	The task was specific.
narrow	The information requested was narrowly focused.
newinfo	The task description provided me with new information that I did not already know about this topic.
figureout	I needed to figure out things that were not specified in the task description.
details	The task description had details that helped me complete the task.
lookfor	I knew some specific things to look for to address the task.
exactitems	The task description told me exact items I needed to compare.
exactdims	The task description told me exact criteria I needed to consider when comparing items.
openended	The task was open-ended.
opinions	The task required gathering information regarding people's feelings, tastes, and/or opinions.
factualinfo	The task required gathering factual information.

Table2

プレで使用されたアンケート項目

実験手法:ポストタスクアンケート-3



簡易版

ユーザー・エンゲージメント尺度:

1. 没頭の度合い (Focused Attention)
2. 使用感 (Perceived Usability)
3. 審美的魅力 (Aesthetic appeal)
4. 報酬 (Reward)

実験手法 : ポストタスクアンケート-3

没頭の度合い	FA1: I lost myself in this search experience.
	FA2: The time I spent searching just slipped away.
	FA3: I was absorbed in the search task.
使用感	PU1: I felt frustrated while doing the search task.
	PU2: My search experience was taxing.
	PU3: I found the search system confusing to use.
審美的魅力	AE1: The search system was attractive.
	AE2: The search system was aesthetically appealing.
	AE3: The search system appealed to my senses.
報酬	RW1: My search experience was worthwhile.
	RW2: My search experience was rewarding.
	RW3: I felt interested in the search task.

Table3
簡易版ユーザー・エンゲージメント尺度

実験手法：データ分析

- 144人 × 6試行=864試行のうち、25試行を不注意、必要事項の欠如から除外
- シャピロウィルク検定の結果、プレ/ポストアンケート、UES-SFのいずれも正規性なし
→log変換をしてANOVAを実施

3.結果

結果：プレアンケート

RQ1. タスクの種類に参加者の**プレアンケートへの影響**はどのようなものか

RQ2. タスクの種類に参加者の**ポストアンケートへの影響**はどのようなものか

RQ3. タスクの種類に参加者の**タスク中のエンゲージメントへの影響**はどのようなものか

RQ4. タスクの種類**の検索行動への影響**はどのようなものか

RQ5. タスクの種類**の検索戦略への影響**はどのようなものか

結果：概要

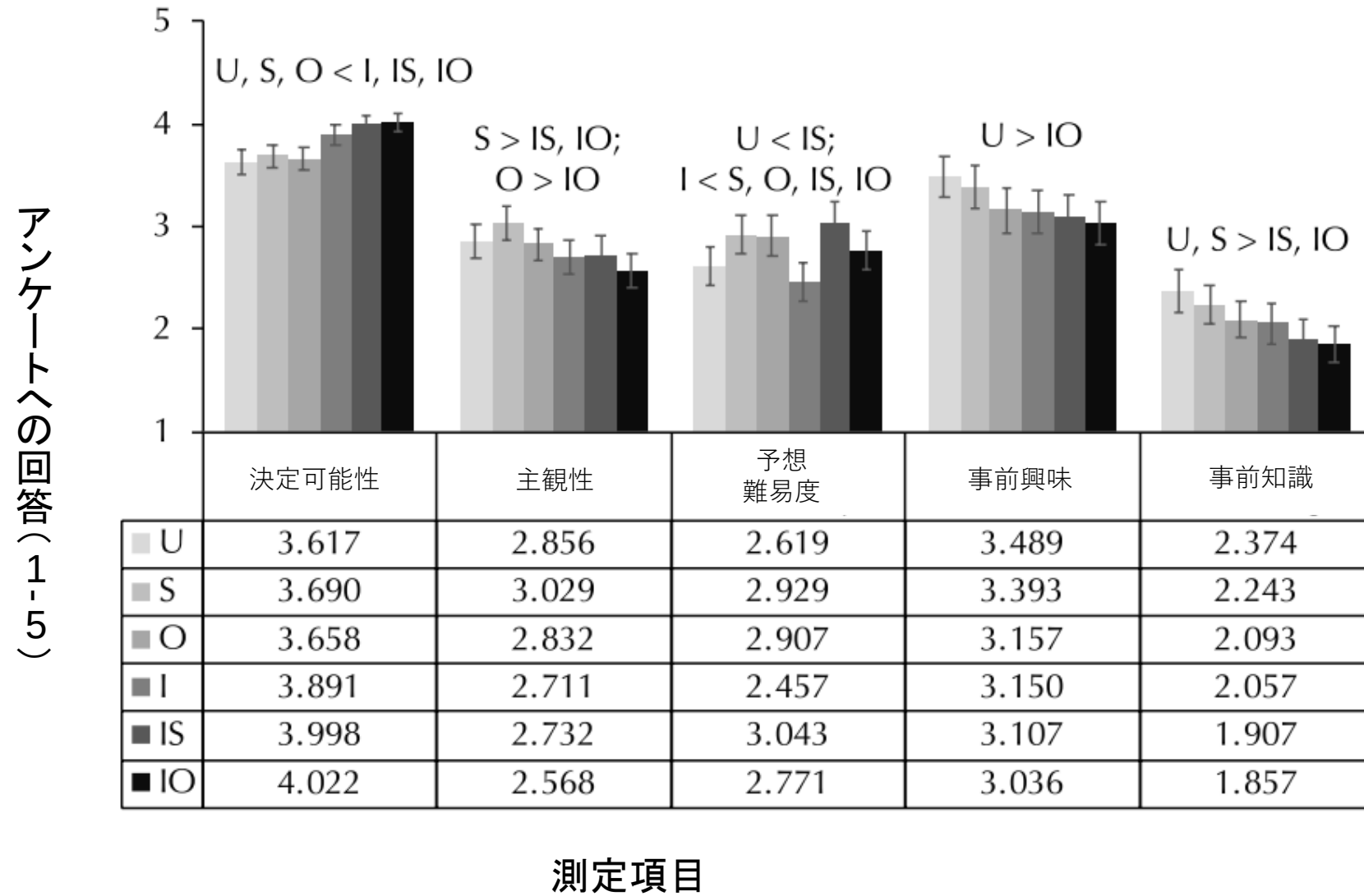
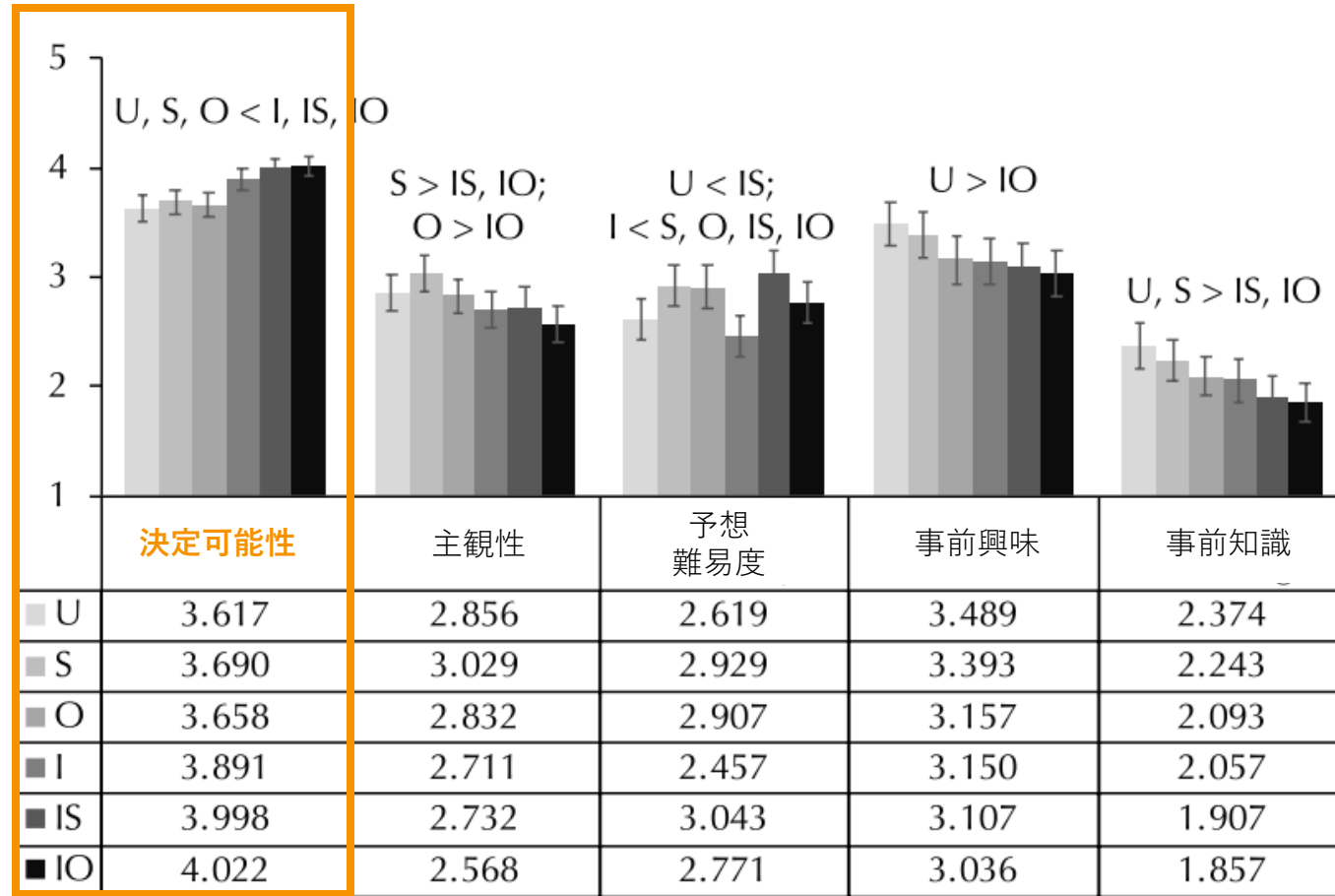


Fig.2 プレアンケートの結果

結果：プレアンケート（決定可能性）

アンケートへの回答（1-5）



測定項目

タスクの種類の主効果あり
 $(F(4.103, 566.171) = 20.431, p = .001)$

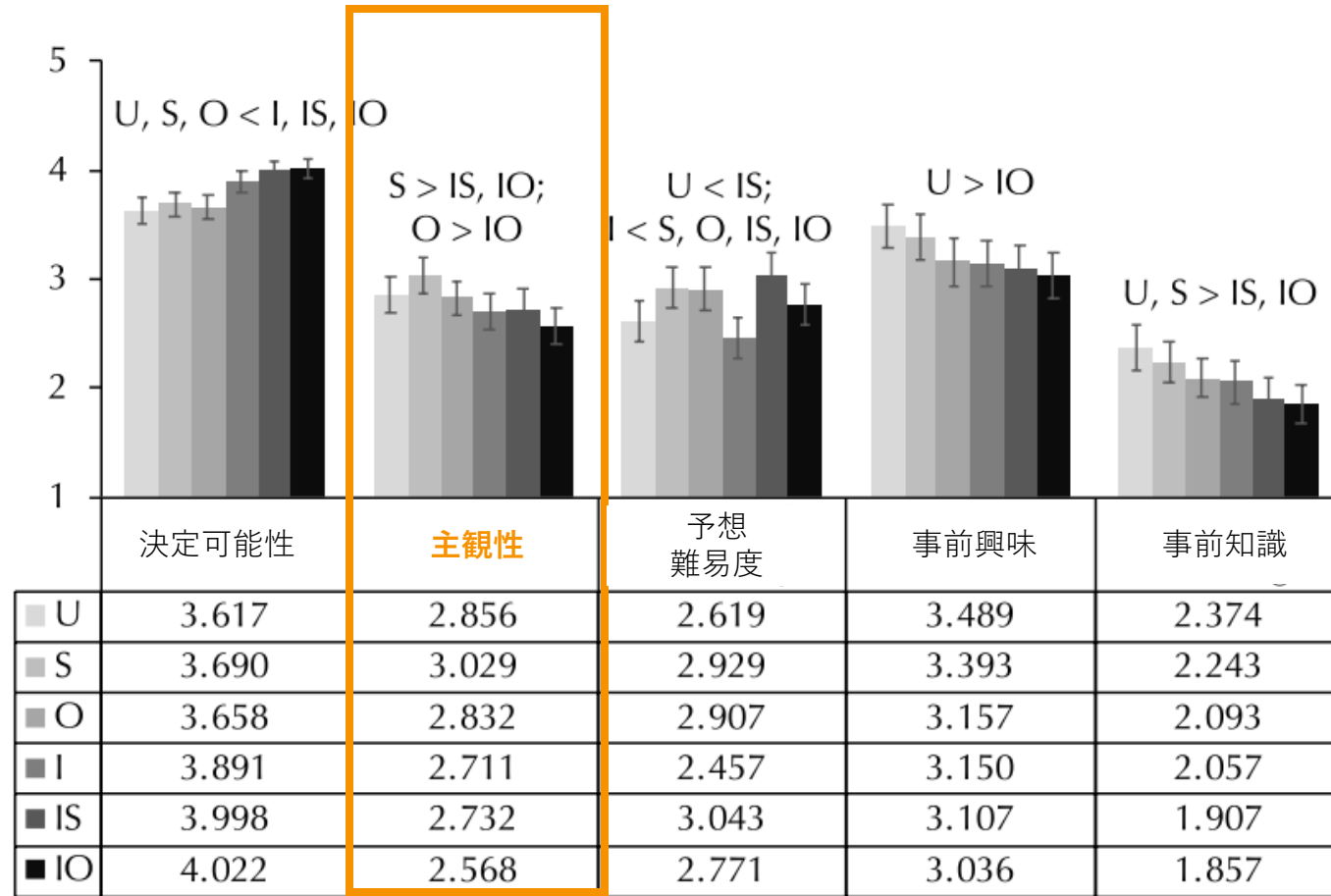
事後検定
 $U, S, O < I, IS, IO$
 $(p < .05)$

**アイテムの限定は
決定可能性を有意に高めた
次元の限定は影響なし**

Fig.2 プレアンケートの結果

結果：プレアンケート(主観性)

アンケートへの回答(1-5)



測定項目

タスクの種類の主効果あり

($F(5, 690) = 6.518, p = .001$)

事後検定

S > IS, IO

O > IO

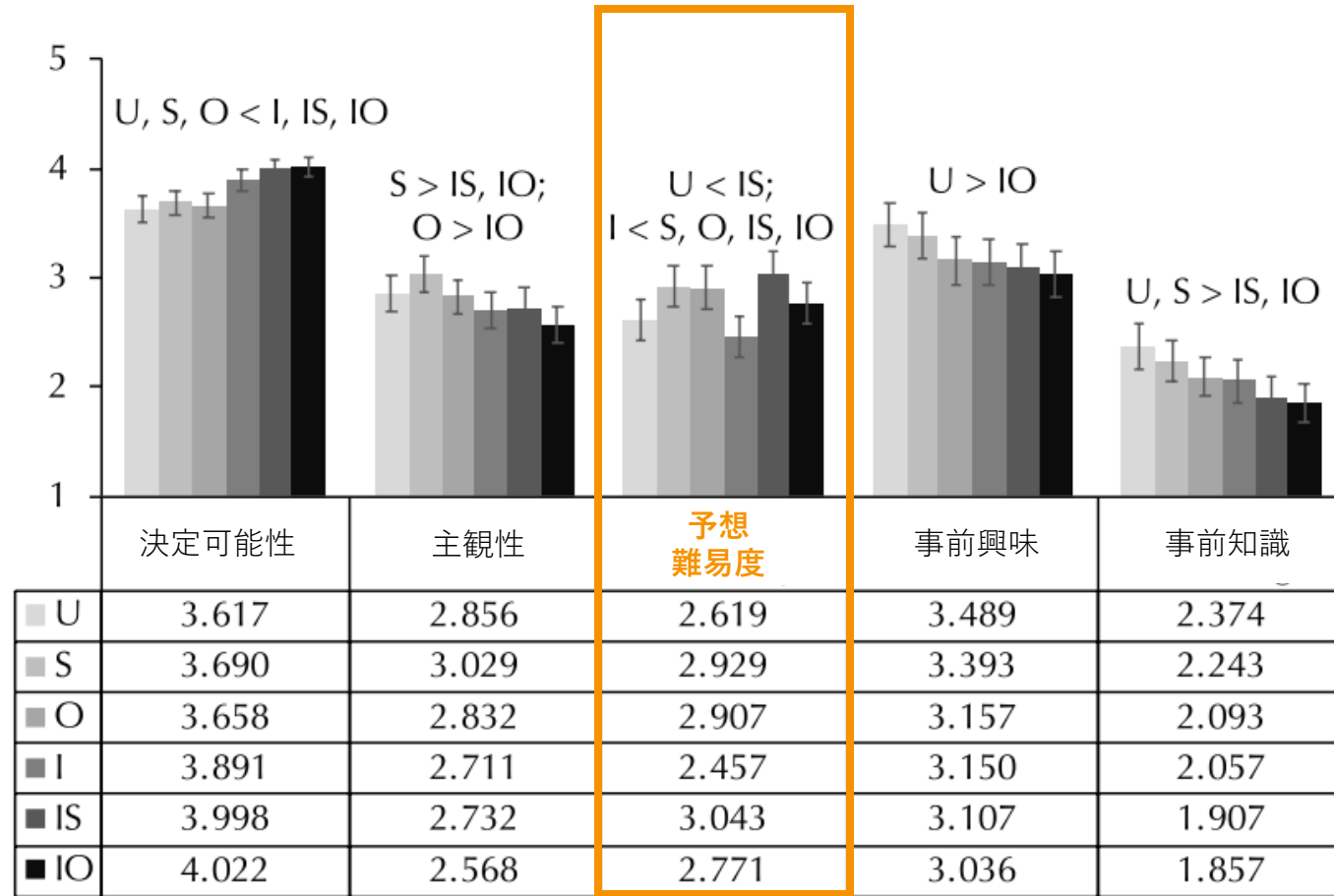
($p < .05$)

アイテムの
タスク範囲からの除外は
有意に主観性を高めた

Fig.2 プレアンケートの結果

結果：プレアンケート（予想難易度）

アンケートへの回答（1-5）



測定項目

タスクの種類の主効果あり

($F(5, 690) = 7.071, p = .001$)

事後検定

$U < IS$

$I < S, O, IS, IO$

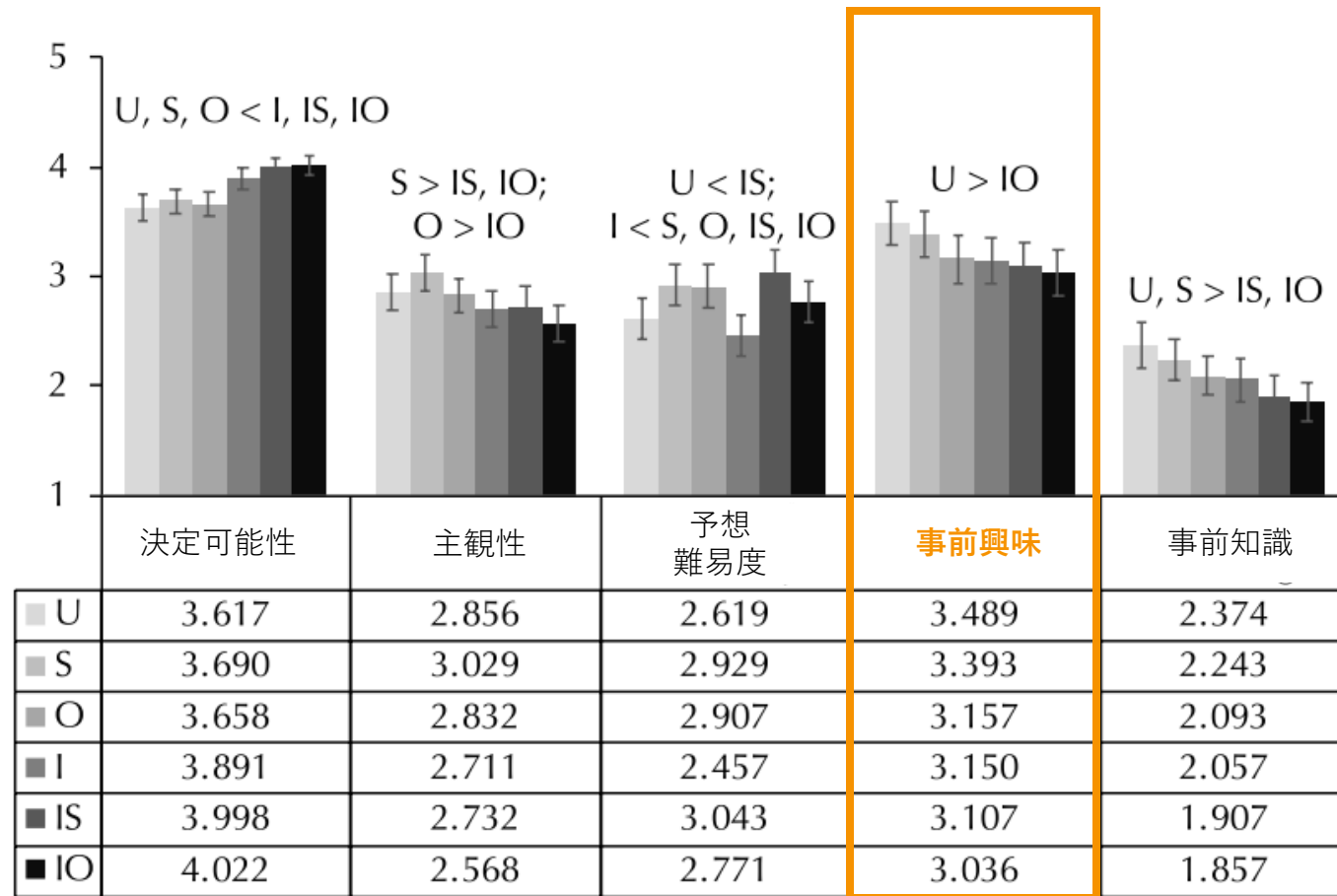
($p < .05$)

次元の限定は、
難易度評価を有意に高めた

Fig.2 プレアンケートの結果

結果：プレアンケート（事前興味）

アンケートへの回答（1-5）



測定項目

タスクの種類の主効果あり

($F(5, 690) = 3.832, p = .002$)

事後検定

U > IO

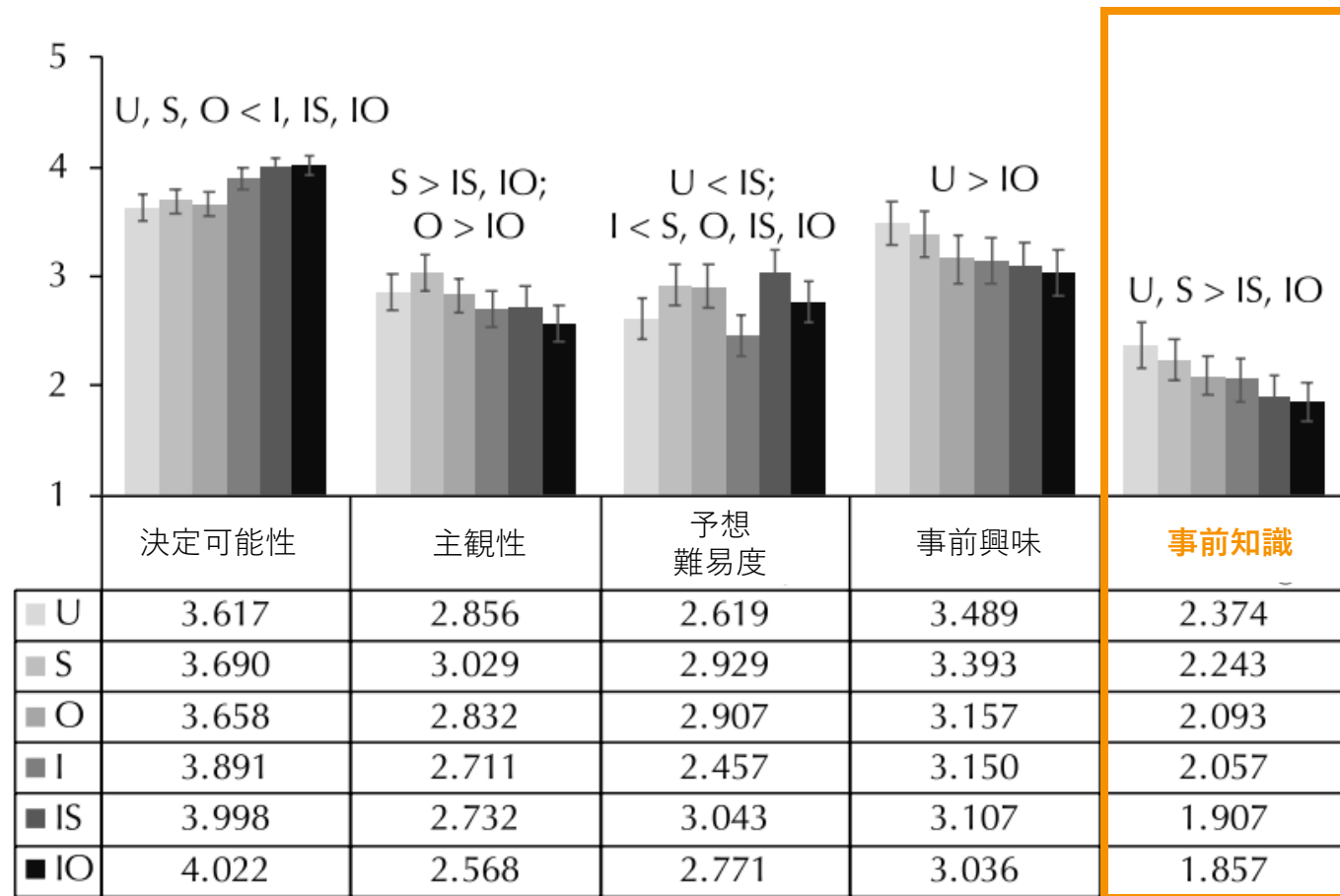
($p < .05$)

アイテム、次元の限定は
事前興味を有意に低めた

Fig.2 プレアンケートの結果

結果：プレアンケート（事前知識）

アンケートへの回答（1-5）



測定項目

タスクの種類の主効果あり

($F(5, 690) = 6.441, p = .000$)

事後検定

U, S > IS, IO
($p < .05$)

アイテム、次元の限定は
事前知識を有意に低めた

Fig.2 プレアンケートの結果

結果：プレアンケート まとめ

- 決定可能性：アイテムの限定は決定可能性を有意に高めたが、次元の限定は影響なし
- 主観性：アイテムのタスク範囲からの除外は有意に主観性を高めた
- 予想難易度：次元の限定は、難易度評価を有意に高めた
- 事前興味：アイテム、次元の限定は事前興味を有意に低めた
- 事前知識：アイテム、次元の限定は事前知識を有意に低めた

結果：ポストアンケート

RQ1. タスクの種類に参加者の**プレアンケート**への影響はどのようなものか

RQ2. タスクの種類に参加者の**ポストアンケート**への影響はどのようなものか

RQ3. タスクの種類に参加者の**タスク中のエンゲージメント**への影響はどのようなものか

RQ4. タスクの種類に**検索行動**への影響はどのようなものか

RQ5. タスクの種類に**検索戦略**への影響はどのようなものか

結果：概要

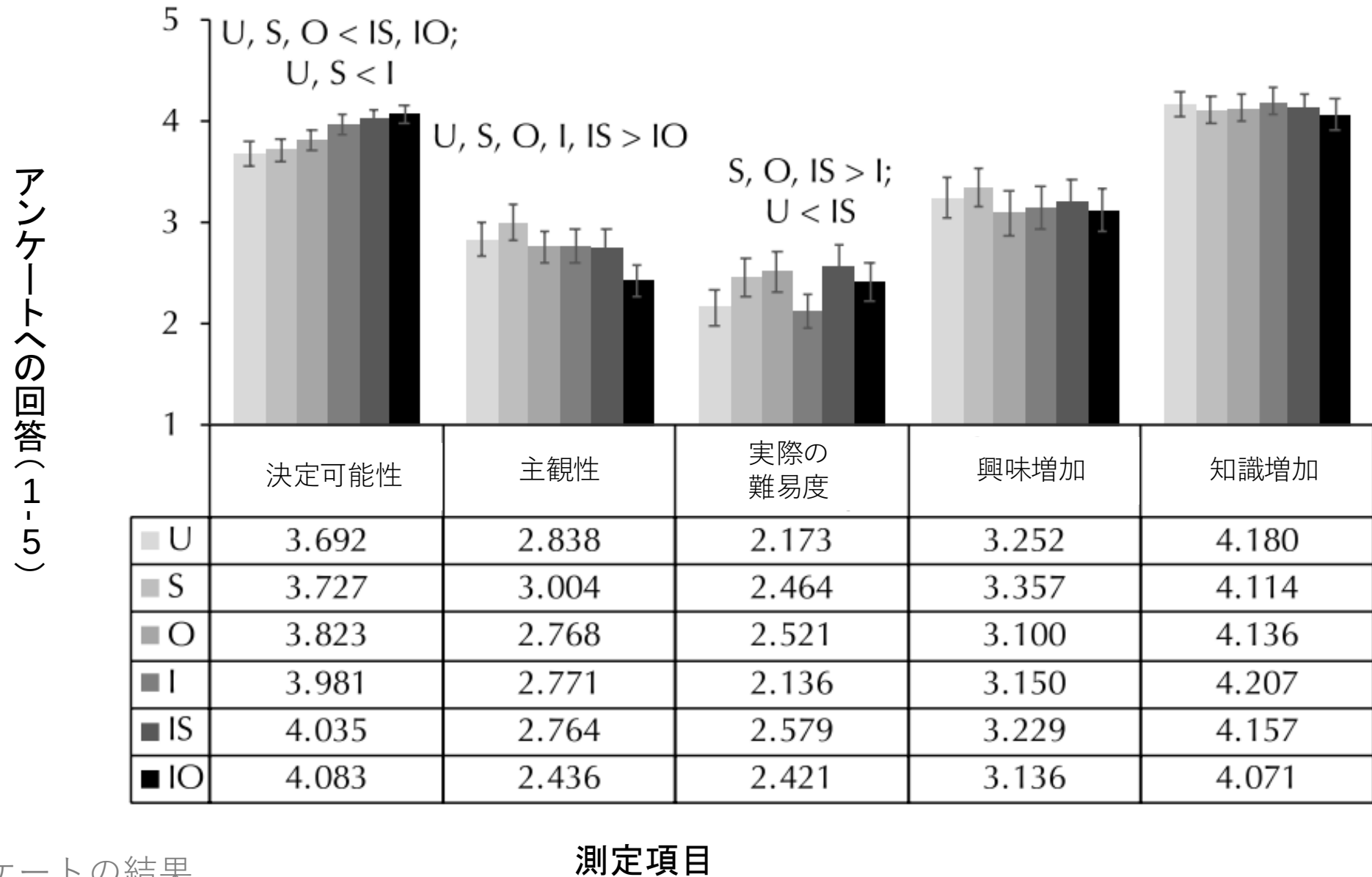
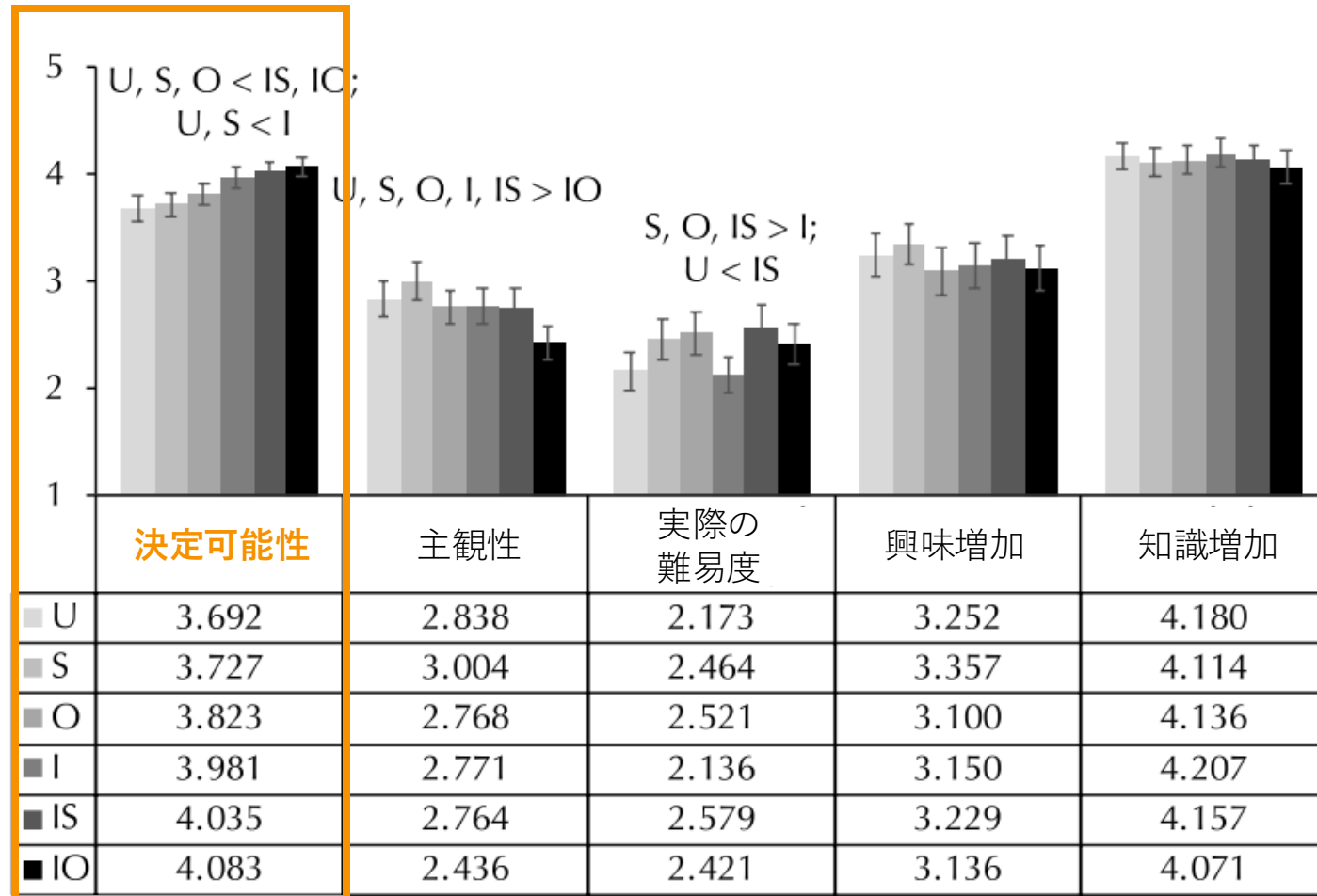


Fig.3 ポストアンケートの結果

結果：ポストアンケート（決定可能性）

アンケートへの回答（1-5）



測定項目

タスクの種類の主効果あり

($F(4.161, 574.201) = 19.484, p = .001$)

事後検定

$U, S, O < \mathbf{IS, IO}$

$U, S < \mathbf{I}$

($p < .05$)

アイテムの限定は

決定可能性を有意に高めたが、

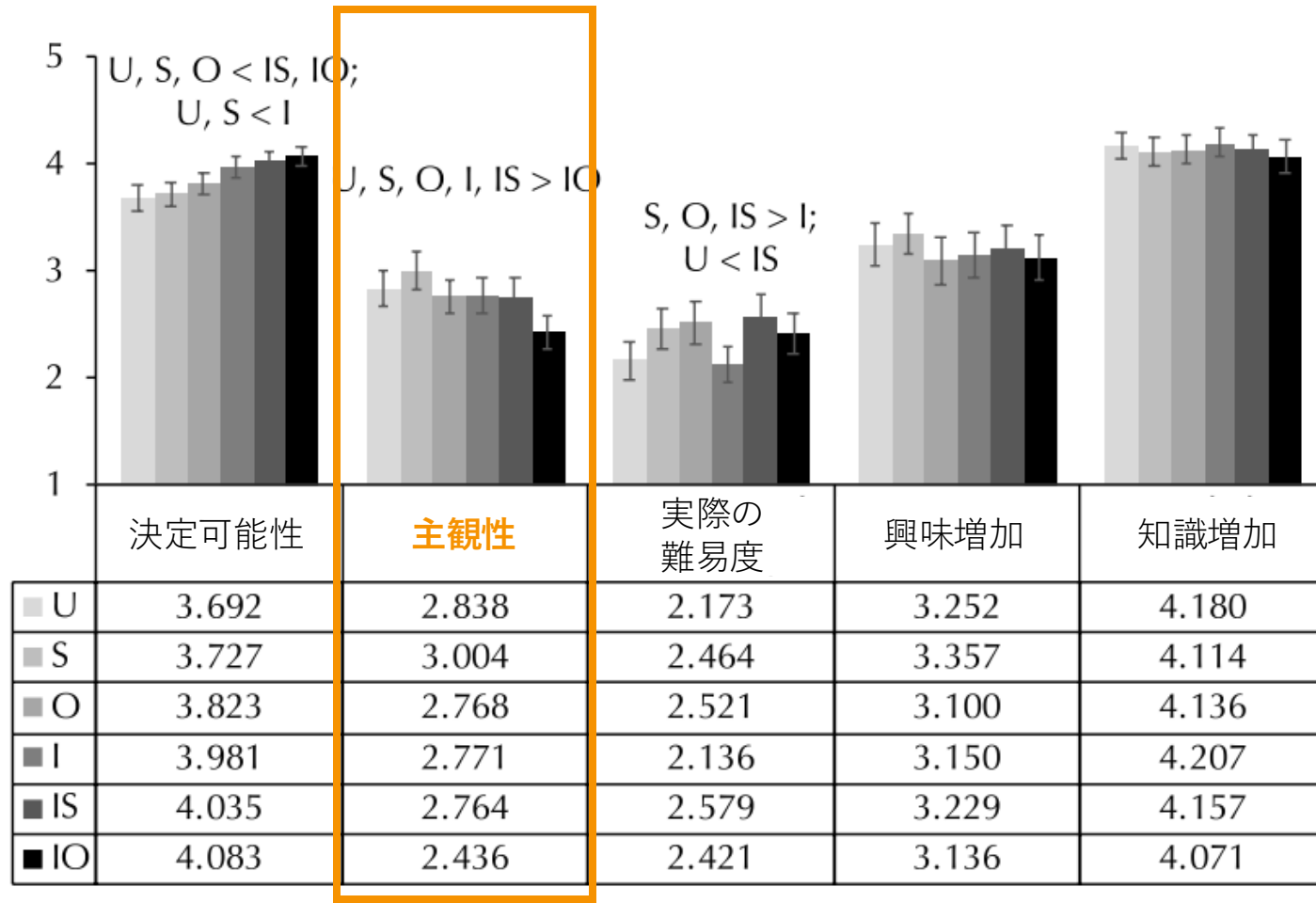
次元の限定は影響なし

*プレと同様

Fig.3 ポストアンケートの結果

結果：ポストアンケート（主観性）

アンケートへの回答（1-5）



測定項目

タスクの種類の主効果あり

($F(5, 690) = 8.877, p = .001$)

事後検定

U, S, O, I, IS > IO

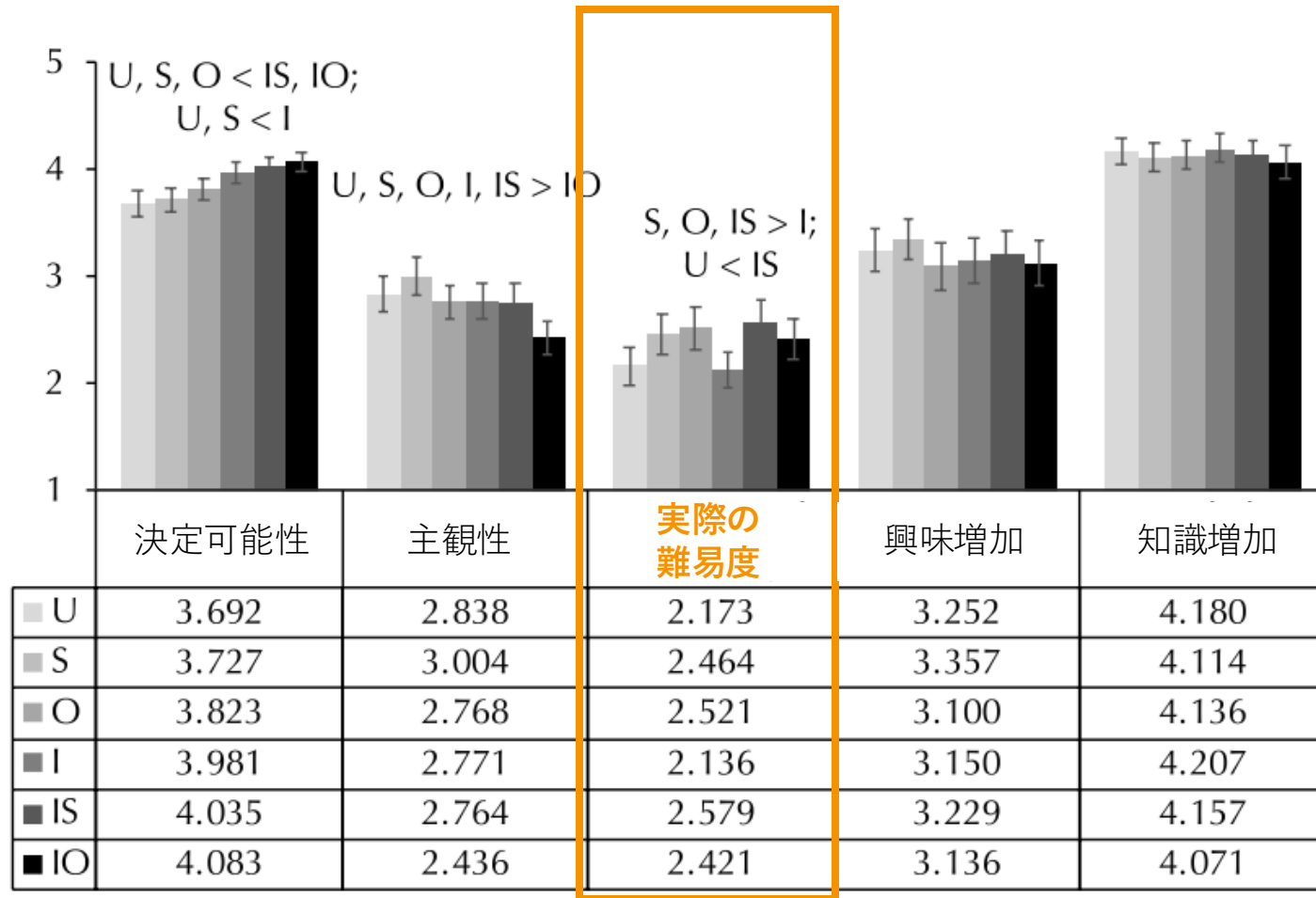
($p < .05$)

**IOタスクにおいて、
他の全てのタスクよりも
主観性を有意に低めた**

Fig.3 ポストアンケートの結果

結果：ポストアンケート（実際の難易度）

アンケートへの回答（1-5）



測定項目

タスクの種類の主効果あり

($F(5, 690) = 4.239, p = .001$)

事後検定

S, O, IS > I

U < IS

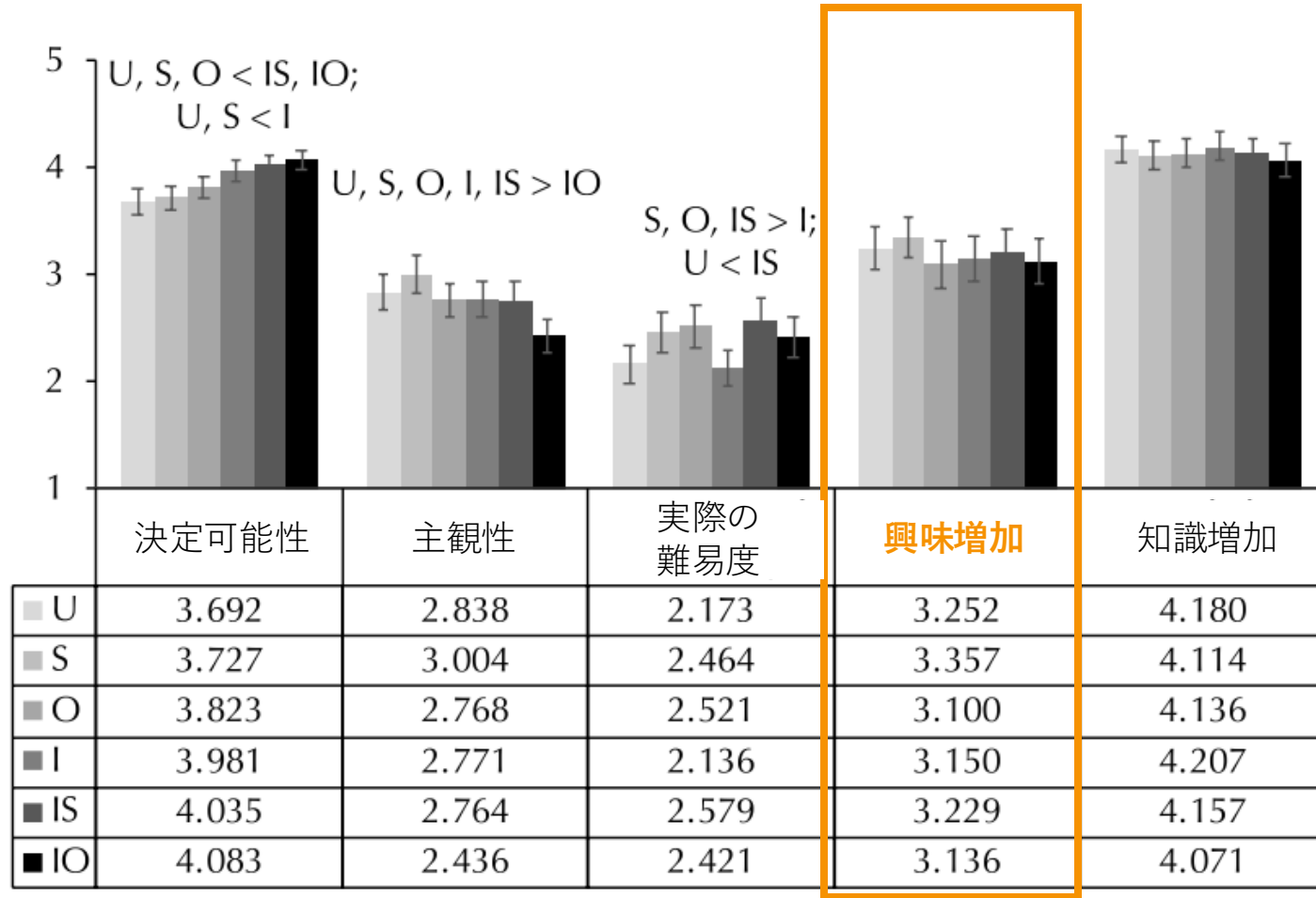
($p < .05$)

次元の限定が
実際の難易度を有意に高めた

Fig.3 ポストアンケートの結果

結果：ポストアンケート（興味増加）

アンケートへの回答（1-5）



タスクの種類の主効果なし
($F(4.611, 636.343) = 1.590, p = .166$)

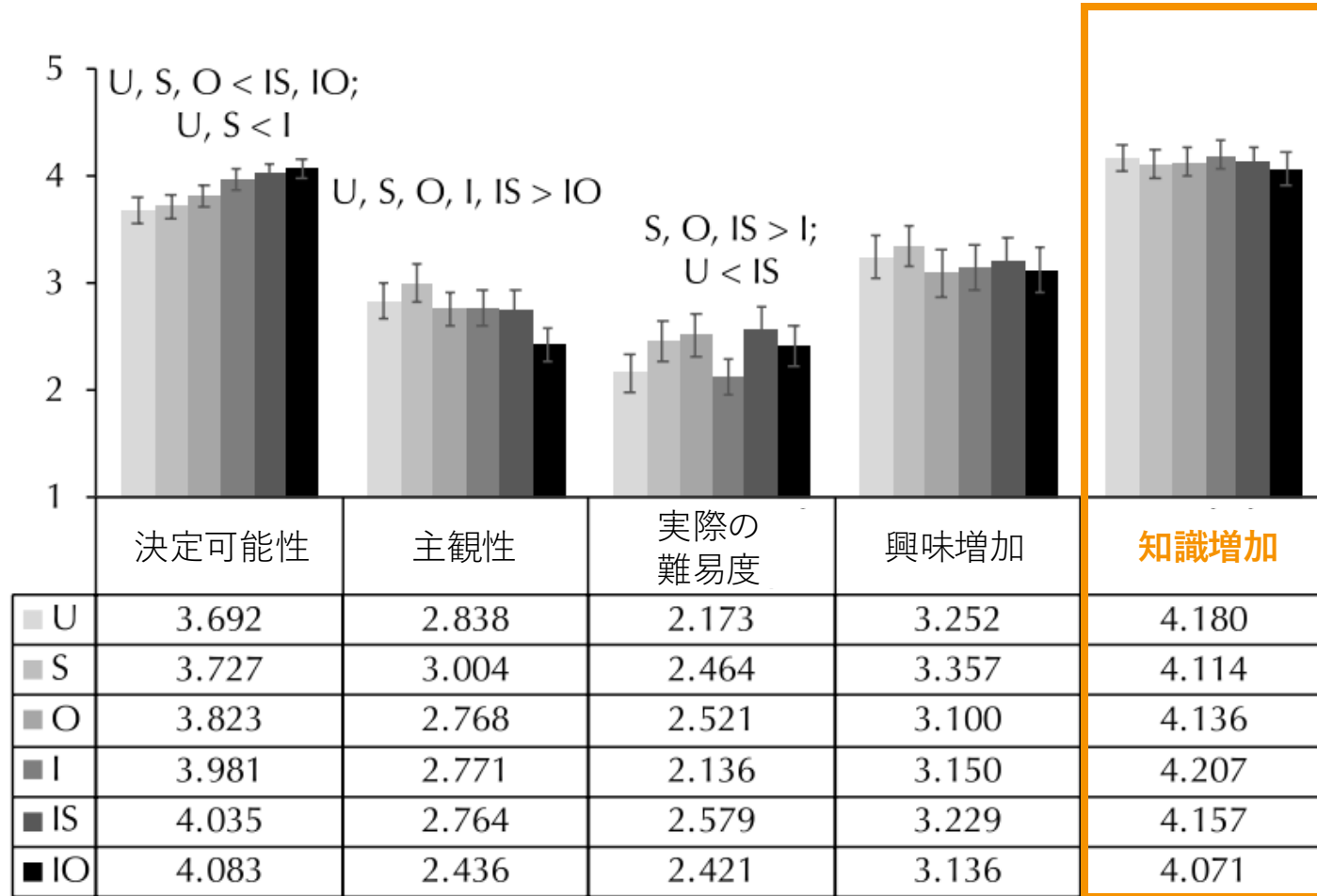
$S > O$
 $IS > IO$
(n.s.)

測定項目

Fig.3 ポストアンケートの結果

結果：ポストアンケート（知識増加）

アンケートへの回答（1-5）



測定項目

タスクの種類の主効果なし
($F(4.478, 617.965) = .919, p = .460$)

全てのタスクにおいて
一貫して高い (Over 4/5)

Fig.3 ポストアンケートの結果

結果：ポストアンケートまとめ

- 決定可能性：**アイテムの限定**は**決定可能性を有意に高めたが、次元の限定は影響なし**
- 主観性：**IOタスク**において、**主観性を有意に低めた**
- 実際の難易度：**次元の限定**が**実際の難易度を有意に高めた**
- 興味増加：タスクの種類の主効果なし
- 知識増加：タスクの種類の主効果なし

結果：ポストアンケート（UES）

RQ1. タスクの種類に参加者の**プレアンケート**への影響はどのようなものか

RQ2. タスクの種類に参加者の**ポストアンケート**への影響はどのようなものか

RQ3. タスクの種類に参加者の**タスク中のエンゲージメント**への**影響**はどのようなものか

RQ4. タスクの種類に**検索行動**への影響はどのようなものか

RQ5. タスクの種類に**検索戦略**への影響はどのようなものか

結果：概観

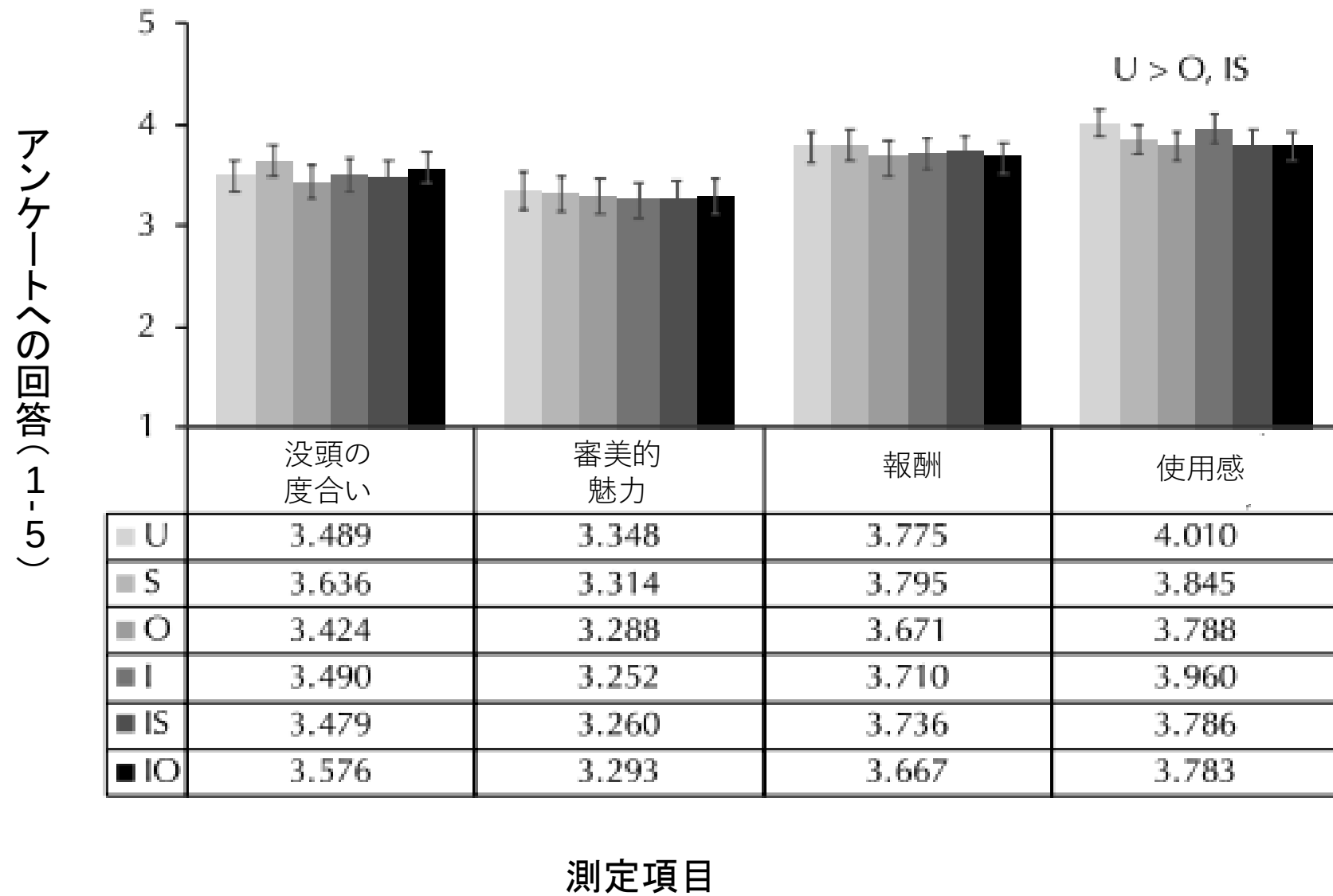
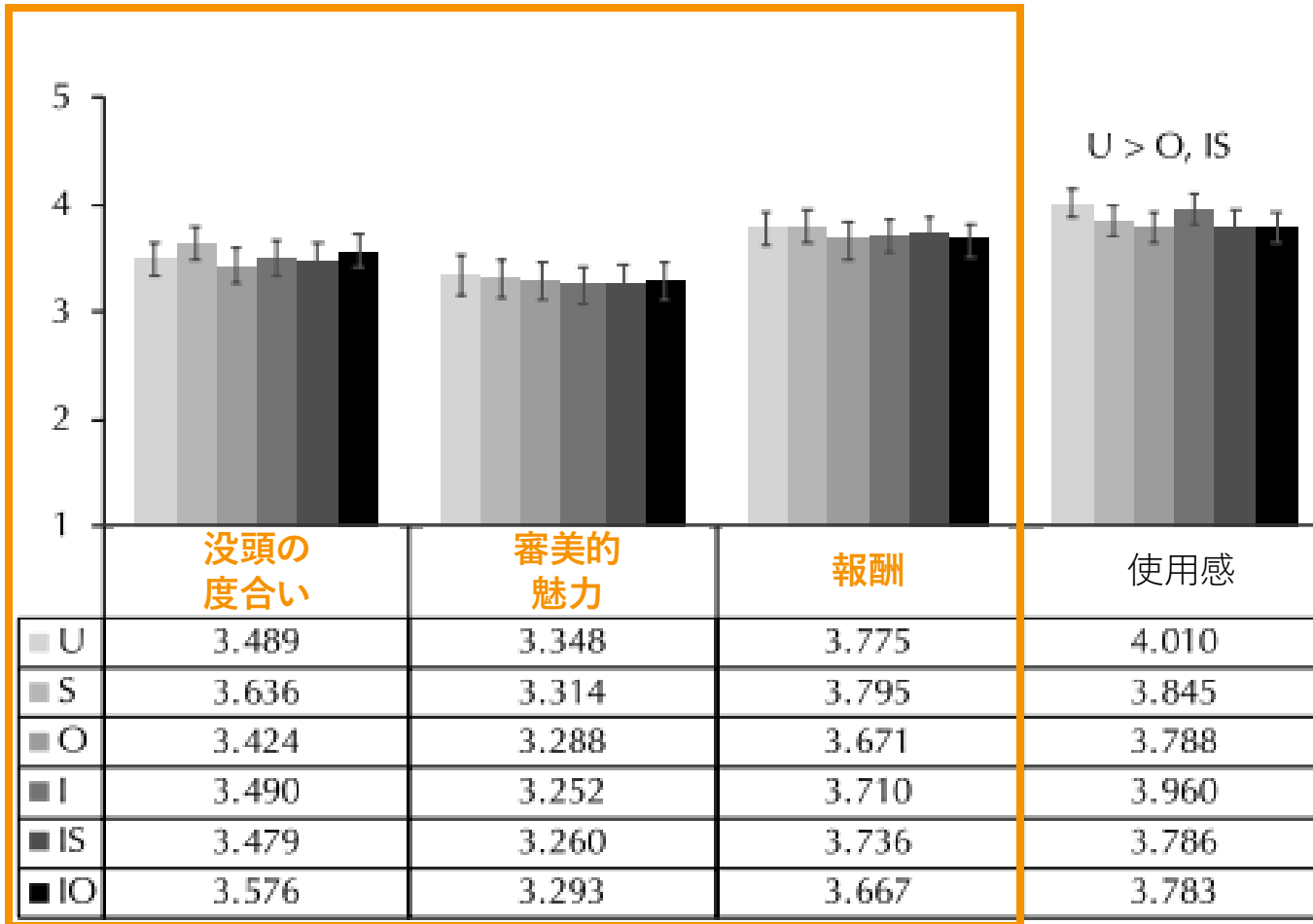


Fig.4 ポストアンケート(UES)の結果

結果：ポストアンケート（UES）

アンケートへの回答（1-5）



測定項目

タスクの種類の主効果なし

没頭の度合い:

$(F(4.495, 620.311) = 2.141, p = .066)$

審美的魅力:

$(F(3.889, 536.707) = .924, p = .447)$

報酬:

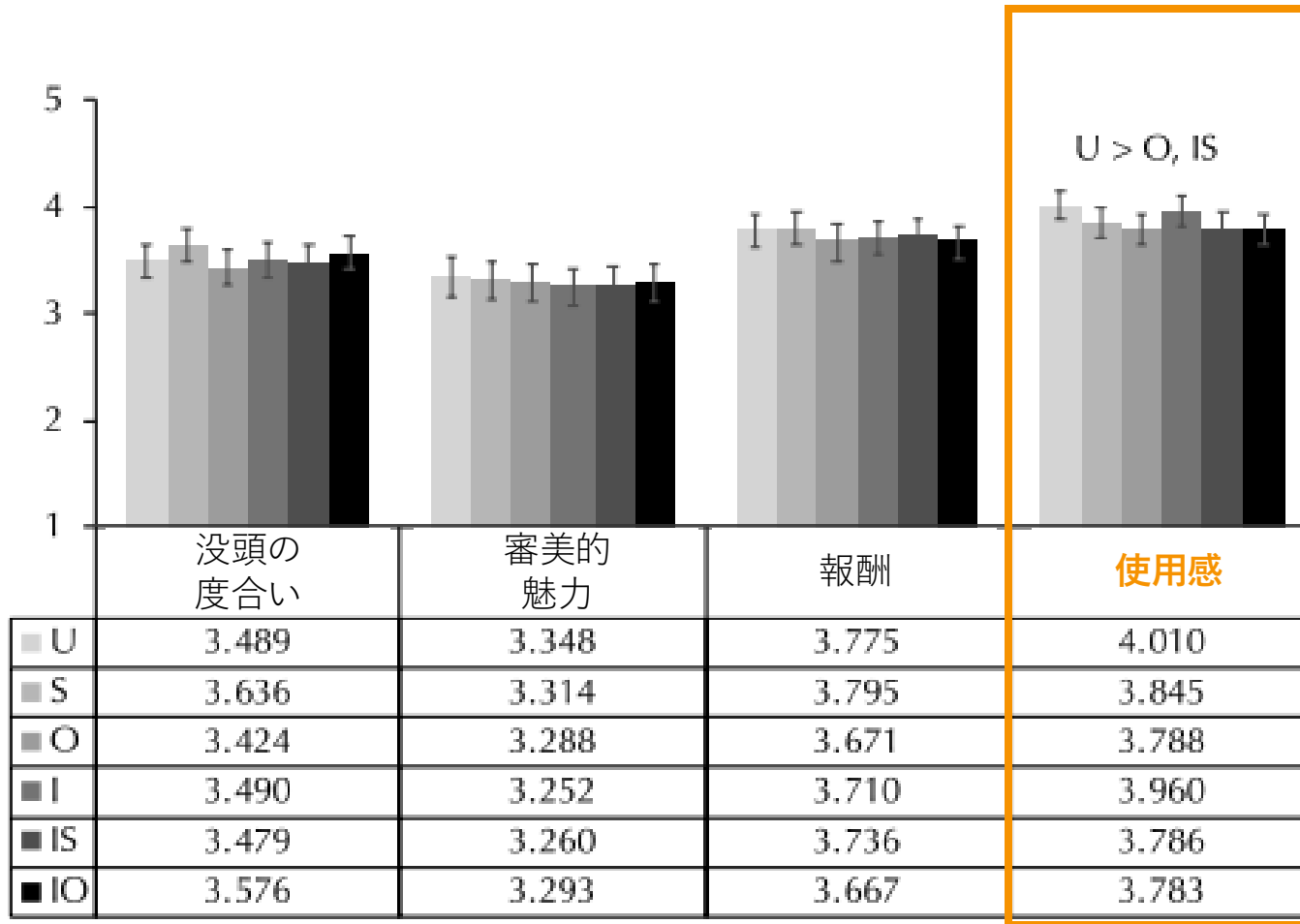
$(F(4.359, 601.524) = 1.017, p = .401)$

U > O, IS

Fig.4 ポストアンケート（UES）の結果

結果：ポストアンケート（UES-使用感）

アンケートへの回答（1-5）



測定項目

タスクの種類の主効果あり

($F(4.601, 634.946) = 3.024, p = .013$)

事後検定

U > O, IS ($p < .05$)

U > IO ($p = .070$)

次元の限定が使用感を低めた

Fig.4 ポストアンケート（UES）の結果

結果：ポストアンケート(UES)まとめ

- 没頭の度合い：タスクの種類の主効果なし
- 審美的魅力：タスクの種類の主効果なし
- 報酬：タスクの種類の主効果なし
- 使用感：**次元の限定が使用感を低めた**

結果：

RQ1. タスクの種類に参加者の**プレアンケート**への影響はどのようなものか

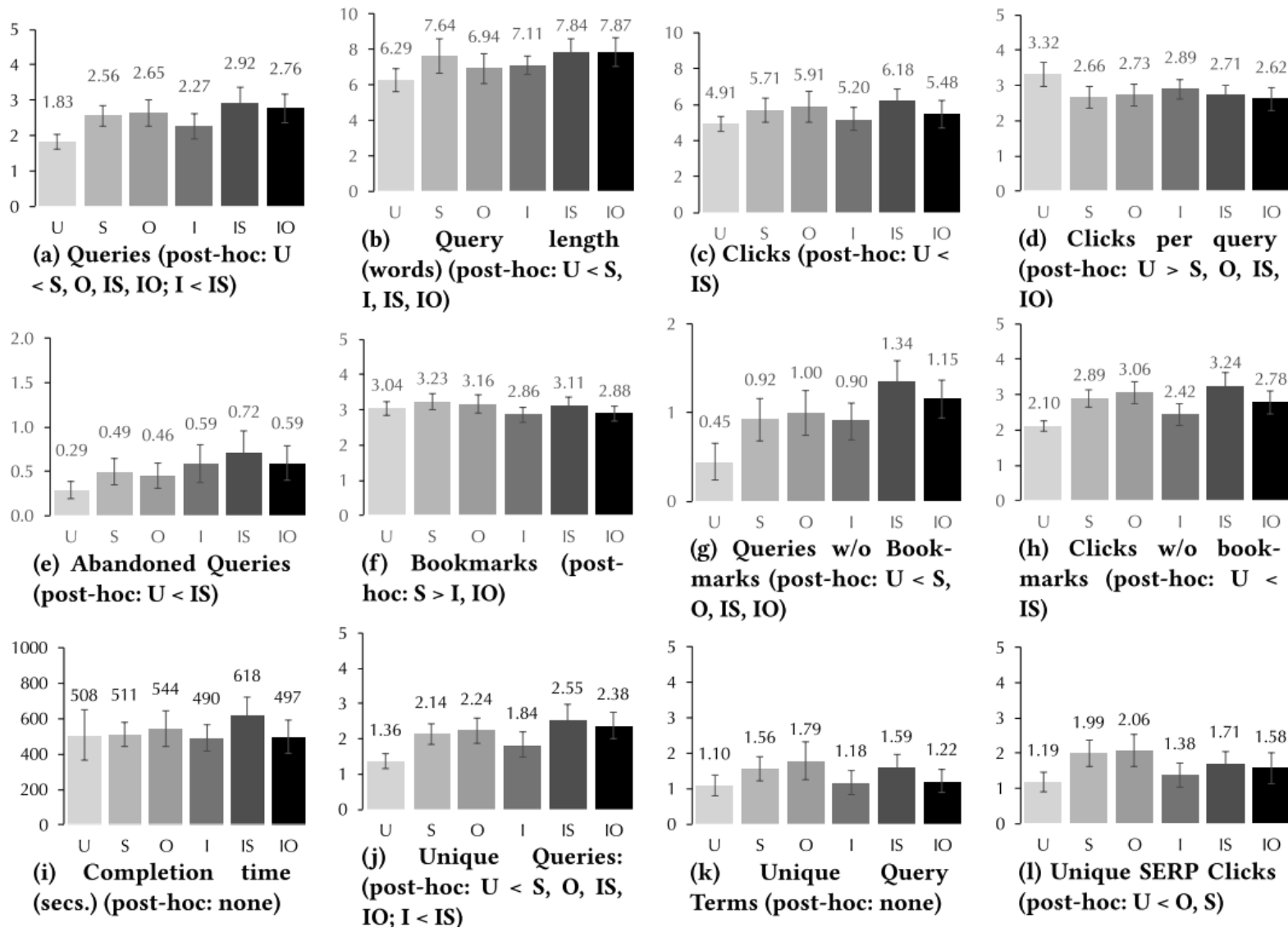
RQ2. タスクの種類に参加者の**ポストアンケート**への影響はどのようなものか

RQ3. タスクの種類に参加者の**タスク中のエンゲージメント**への影響はどのようなものか

RQ4. タスクの種類を検索行動への影響はどのようなものか

RQ5. タスクの種類**の検索戦略への影響**はどのようなものか

結果：概要



主効果あり

- (a) queries ($F(4.254, 587.120) = 6.519, p < .001$)
- (b) query length ($F(4.412, 608.792) = 6.608, p < .001$)
- (c) clicks ($F(3.976, 548.649) = 2.824, p = .025$)
- (d) clicks per query ($F(5, 690) = 3.755, p = .002$)
- (e) abandoned queries ($F(4.002, 552.276) = 3.162, p = .014$)
- (f) bookmarks ($F(4.658, 642.756) = 3.527, p = .005$)
- (g) queries w/o a bookmark ($F(4.093, 564.883) = 4.858, p = .001$)
- (h) clicks w/o a bookmark ($F(3.756, 518.385) = 2.836, p = .027$)
- (j) unique queries ($F(4.273, 589.728) = 7.856, p = .000$)
- (k) unique query terms ($F(3.879, 535.261) = 2.985, p = .020$)
- (l) unique SERP clicks ($F(4.154, 573.279) = 4.515, p = .001$)

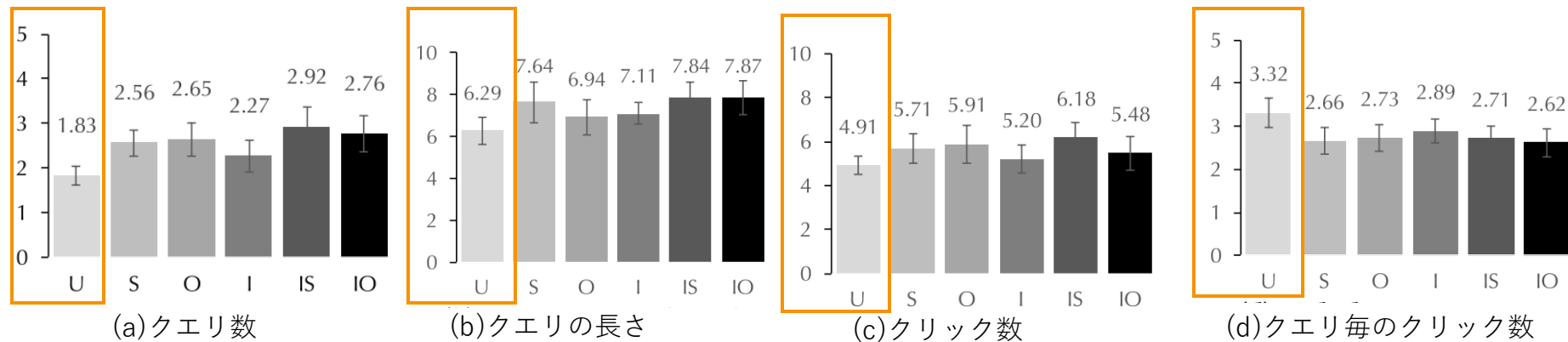
主効果なし

- (i) completion time ($F(3.757, 518.404) = 1.102, p = .354$)

Figure 5

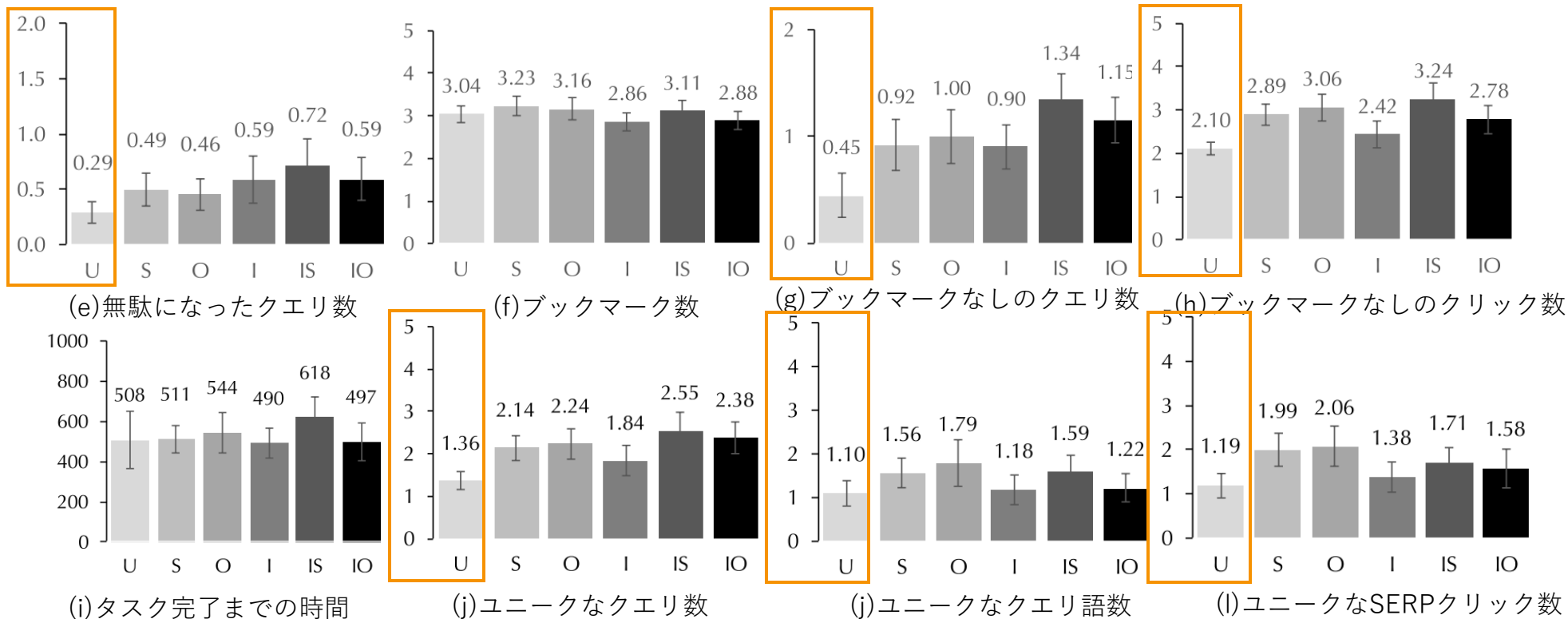
タスクの種類の
検索行動への影響に関する結果

傾向1：U条件が、必要な検索努力が最も少なかった



a, b, c, d, e, g, h, j, k, lにて
最も検索努力の少ない結果

a: $U < S, O, IS, IO; I < IS$
 b: $U < S, I, IS, IO$
 c: $U < IS$
 d: $U > S, O, IS, IO$
 e: $U < IS$
 f: $S > I, IO$
 g: $U < S, O, IS, IO$
 h: $U < IS$
 i: none
 j: $U < S, O, IS, IO; I < IS$
 k: none
 l: $U < O, S$



結果：

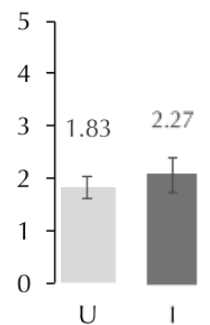
傾向1 考察

U条件において、必要な検索努力が最も少なかった

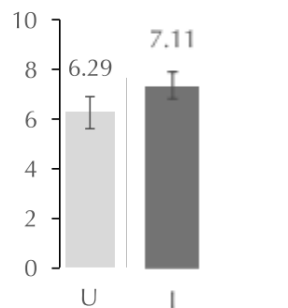


被験者が**Uタスクの完了に対して満足し、
特定のアイテムや次元に関する情報を収集しなかった可能性**

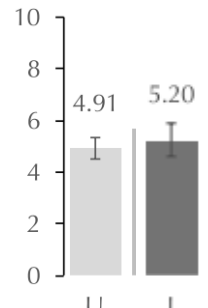
傾向2：I条件の検索努力、ユニークな行動が少なかった



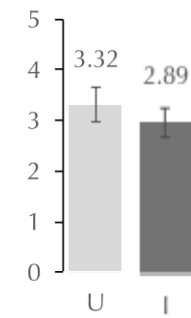
(a) クエリ数



(b) クエリの長さ



(c) クリック数



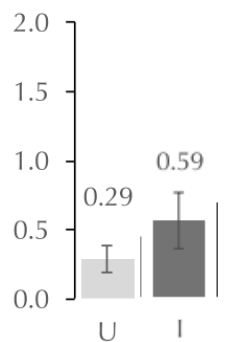
(d) クエリ毎のクリック数

U-I間

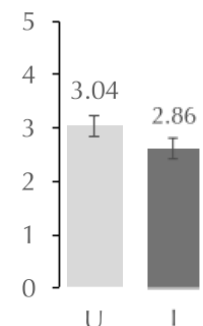
全指標においてn.s.



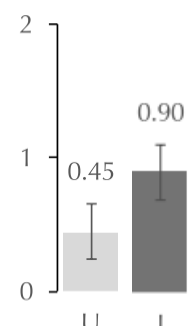
最も検索努力の少ないタスクと
差がないため、
検索努力が少ないと判断



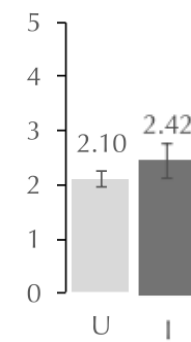
(e) 無駄になったクエリ数



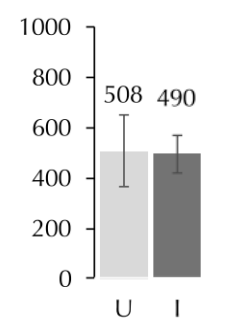
(f) ブックマーク数



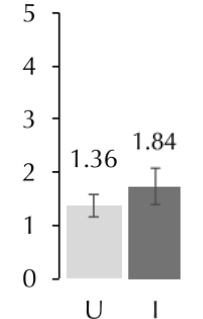
(g) ブックマークなしのクエリ数



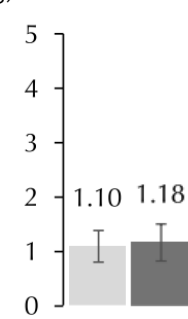
(h) ブックマークなしのクリック数



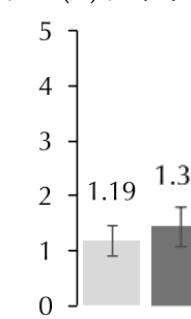
(i) タスク完了までの時間



(j) ユニークなクエリ数



(k) ユニークなクエリ語数



(l) ユニークなSERPクリック数

結果：

傾向2 考察

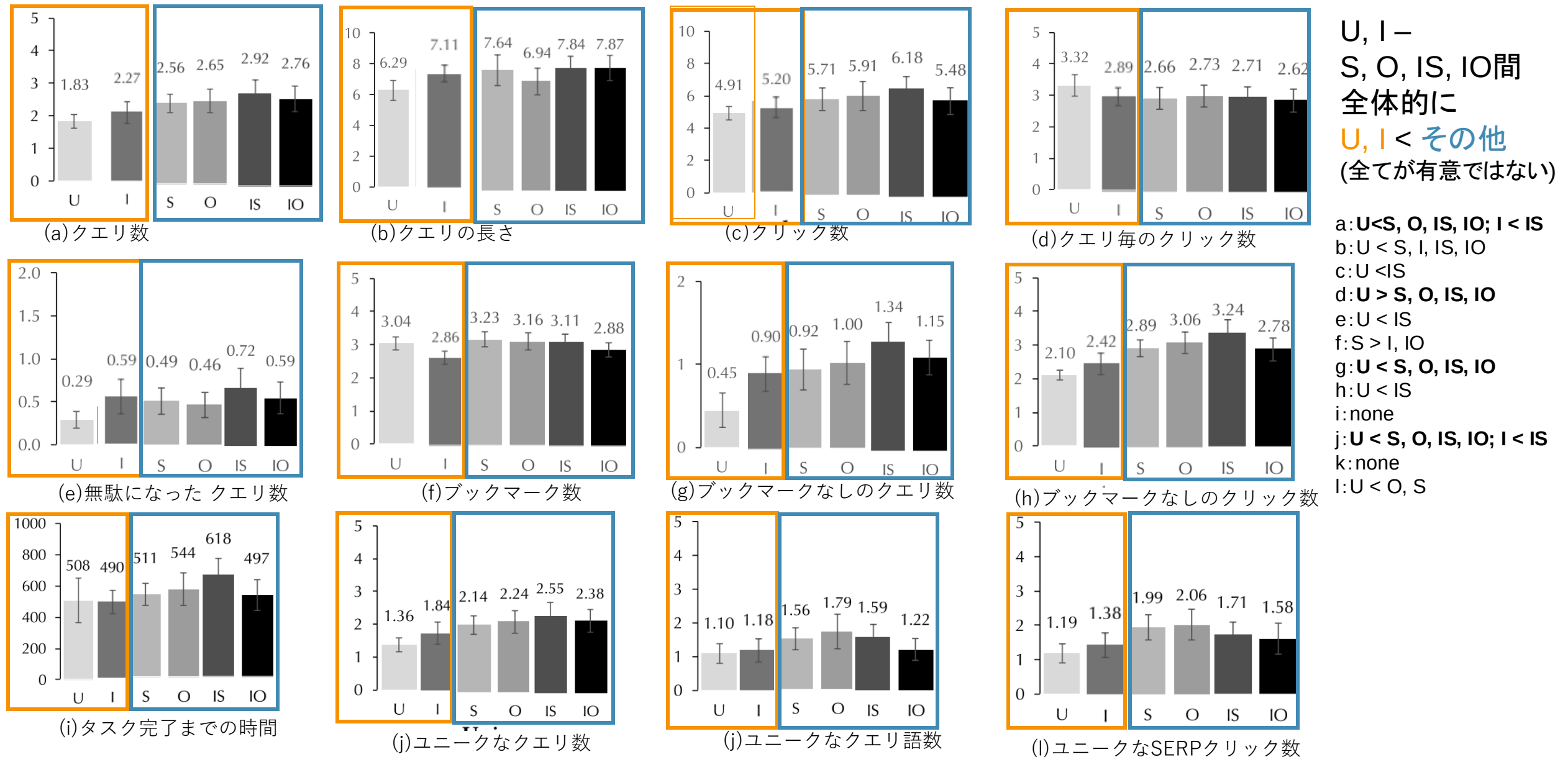
I条件の検索努力、ユニークな行動が少なかった

↓

I条件では、

クエリに含めやすい具体的なアイテムが検索される傾向が強く、
関連するWebサイトに到達しやすかった可能性

傾向3：次元の限定はより大きな検索努力と拡散的な戦略に繋がる



結果：③次元の限定はより大きな検索努力と拡散的な戦略に繋がる

傾向3 考察

次元の限定はより大きな検索努力と拡散的な戦略に繋がる

↓

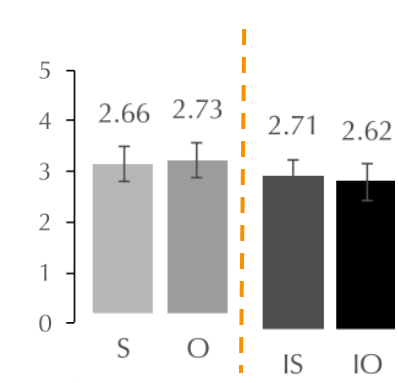
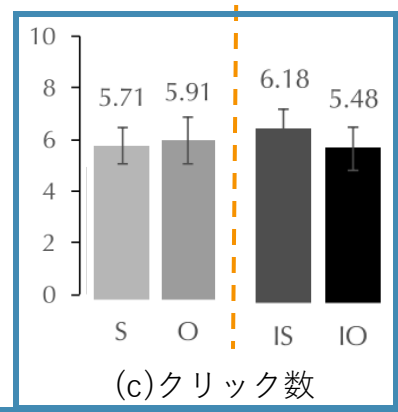
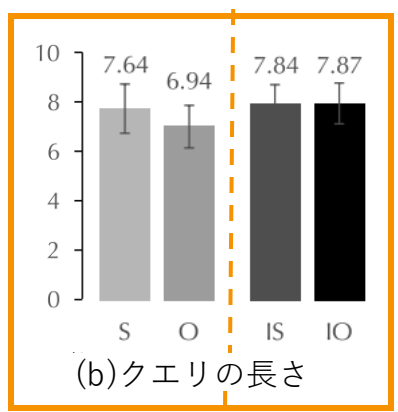
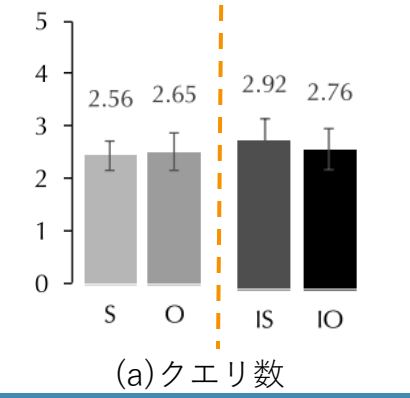
次元の限定は、

複数の言葉で表すことができる抽象的な概念であることから、

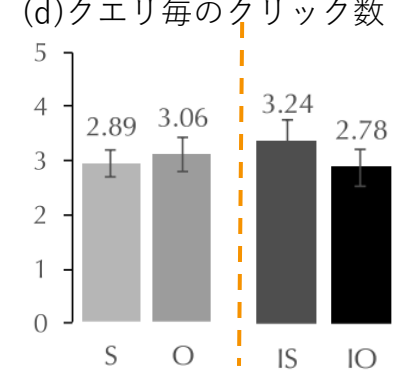
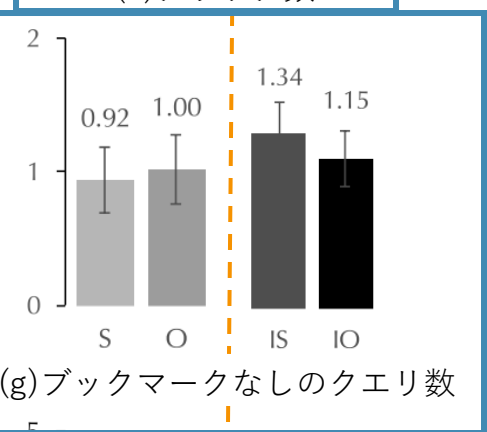
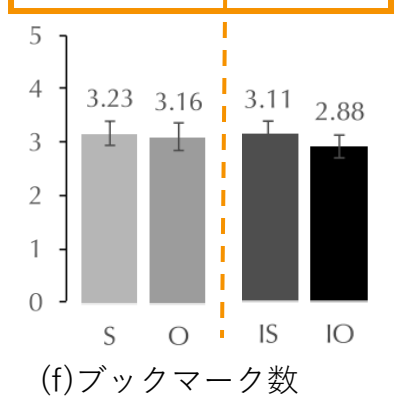
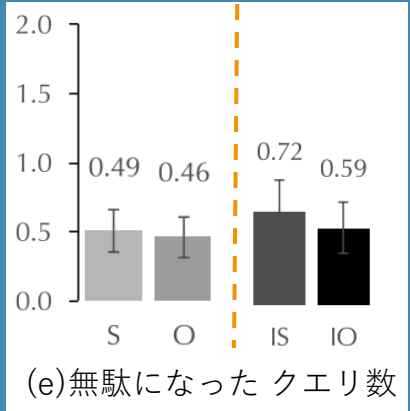
クエリでの表現が難しく、

関連するWebサイトへの到達が難しかった可能性

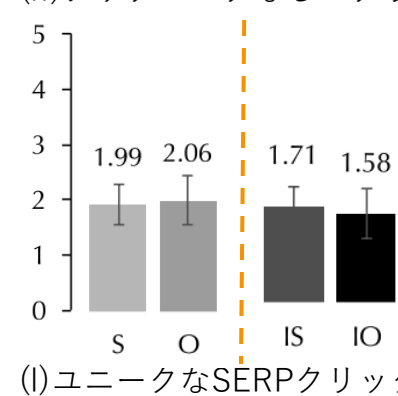
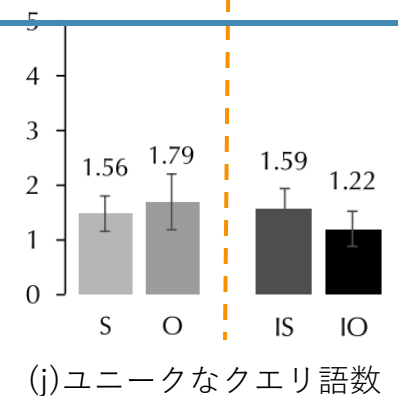
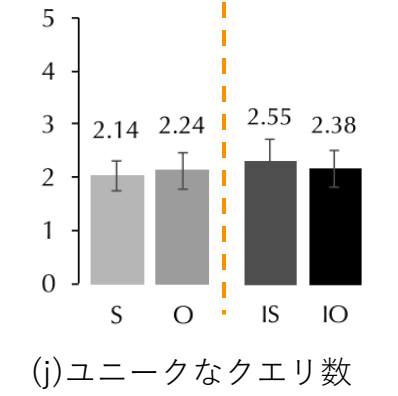
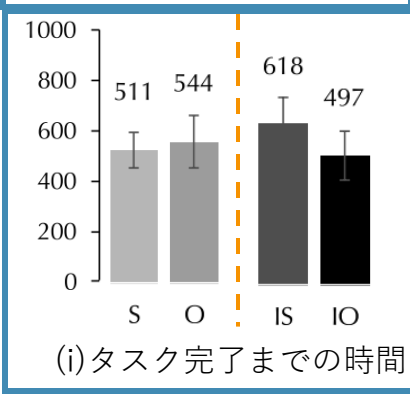
傾向4：主観的な次元有タスクにおいてより多くの検索努力、拡散的な戦略が見られた



S – O間
bにおいて
S > O(n.s.)



IS – IO間
c, e, g, iにおいて
IS > IO (n.s.)



結果：

④主観的な次元有タスクにおいてより多くの検索努力、拡散的な戦略が見られた

傾向4 考察

主観的な次元有タスクにおいて

より多くの検索努力、拡散的な戦略が見られた

↓

主観的な次元は、

より多くの異なるソースからの情報収集、情報統合、

そして情報の信頼性判断を要した可能性がある

結果：検索行動 傾向まとめ

- **U条件**において、**必要な検索努力が最も少なかった**
- **I条件の検索努力、ユニークな行動が少なかった**
- 次元の限定はより大きな検索努力と拡散的な戦略に繋がる
- 主観的な次元有タスクにおいて
より多くの検索努力、拡散的な戦略が見られた

結果：

RQ1. タスクの種類に参加者の**プレアンケート**への影響はどのようなものか

RQ2. タスクの種類に参加者の**ポストアンケート**への影響はどのようなものか

RQ3. タスクの種類に参加者の**タスク中のエンゲージメント**への影響はどのようなものか

RQ4. タスクの種類**の検索行動**への影響はどのようなものか

RQ5. タスクの種類**の検索戦略**への影響はどのようなものか

結果：

実施目的：

1. タスクの種類が、**クエリの生成**に影響したか
2. タスクの種類が、参加者のタスク 1試行実施中の**複数のタイプのクエリの生成**に影響したか

分析対象：

参加者によって生成された全てのクエリ

検索戦略：

クエリに含めるアイテム/項目の有無

結果：コーディング

アイテム

I：一つ以上のアイテムを含む場合

*：それ以外

次元

D：一つ以上の次元を含む場合

*：それ以外

**：最も広い意図

I*, *D：

比較的狭い意図

ID：最も狭い意図

と解釈

結果：コーディング

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

Table4

コーディングの例

結果：コーディングの手法

2人の著者が全体の10%のクエリをそれぞれコーディング

↓

アイテムに関して、Cohen's $k = .966$

次元に関してCohen's $k = .983$

と高水準

↓

残りの90%を2分割してコーディングを実施

結果：タスクの種類が、クエリの生成に影響したか

	(*,*)	(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)

Table5

クエリタイプの分布

オレンジ: 各行の1番多かった戦略

青: 各行において2番目に多かった戦略

結果：タスクの種類が複数のタイプのクエリの生成に影響したか

	$n=1$	$n=2$	$n=3$	$n=4$
U	89.21%	10.07%	0.72%	0.00%
S	48.57%	47.14%	4.29%	0.00%
O	50.36%	41.73%	7.19%	0.72%
I	84.29%	14.29%	0.71%	0.71%
IS	59.29%	37.14%	2.86%	0.71%
IO	64.29%	30.71%	5.00%	0.00%

タスクの種類によって異なる傾向

U, I:

戦略をほとんど変更しなかった

S, O, IS, IO:

戦略の変更が一般的

Table6
1試行内で、 n 種類のクエリを生成した参加者の%

オレンジ: 戦略を変更しなかった人の%

青: 戦略を変更した人の%

結果：まとめ

- Uタスクは決定可能性が最も低く評価されたが、検索努力が最も少ない
- アイテムの限定により、決定可能性の評価は高まり、検索努力も少なくした
- 次元の限定により、決定可能性の評価は高まらず、検索努力は増加した
- 主観的次元の方が、客観的次元よりも多くの検索努力を要した
- タスクの種類はクエリ生成の戦略に影響した
 - U, Iタスクにおいては戦略の変更はほとんどされなかったが、それ以外では戦略の変更が一般的だった

4. ディスカッション

ディスカッション：Uタスク

- プレ/ポストともに最も低い決定可能性だったが、最も検索努力が少なかった
- 戦略として、特定のアイテムや次元のクエリに含めなかった、全てをカバーしようとしなかったのではないか

↓

Satisficing^{*1}や**タスク範囲を限定**することで、
不確実性を低減しながらタスクに臨んでいた可能性

*1 Satisficing:

目的達成のために必要最低限の条件を満たす手順を決定し、追求する行動

ディスカッション：アイテムを限定することの効果

U→I, O→IO, S→ISに関して

- アイテムの限定により、決定可能性が高く評価された
- プレ/ポストの難易度、検索努力に関してU-I間に有意な差がなかった

↓

アイテムの検索は簡単だった可能性

↓

アイテムはクエリに含みやすく、

検索結果を減らすことと検索の過程の不確実性の低減により、決定可能性の増加につながった

ディスカッション：次元を限定することの効果

$U \rightarrow O, U \rightarrow S, I \rightarrow IO, I \rightarrow IS$

- 次元の限定も、アイテムの限定と同様にタスクの範囲を狭めたが、参加者は決定可能性を高く評価せず、より難易度が高いと知覚した

↓

次元に対してのクエリは難しかった可能性

↓

次元は抽象的な概念であり、

様々な方法で表現されることが多いことから、クエリに次元を含めて**関連するWebサイトとを見つけることが困難だった**のではないかと

ディスカッション：エンゲージメントへの効果

- 今回の実験はエンゲージメントへの強い効果は見られなかった

↓

タスクが簡単/難しすぎて、
飽き/フラストレーションや意欲の低下(disengagement)を生じさせた可能性

- タスクの範囲を狭めていたタスクに対してプレの事前興味とポストの使用感を高く評価していた

↓

オープンエンドのタスクは参加者自身の興味を追求を可能にする可能性

ディスカッション：客観条件 vs 主観条件

- 客観条件と主観条件に大きな差は見られなかった
 - 全ての変数間に有意な差なし
- S, IS条件でO, IO条件よりも高い主観性を報告(n.s.)
- IS条件で、IO条件よりも検索努力を要した(n.s.)

↓

主観的な情報は

主観性を高めることを参加者が認識していた可能性

主観的な次元は

異なる視点を要し、検索努力がより必要となった可能性

参考文献

- Byström, K., & Järvelin, K. (1995). Task complexity affects information seeking and use. *Information Processing & Management*, 31(2), 191–213.
- Bell, D. J., & Ruthven, I. (2004). Searchers' assessments of task complexity for web searching. In *Proceedings of the European Conference on Information Retrieval (ECIR)* (pp. 57–71). Springer.
- Capra, R., Arguello, J., Crescenzi, A., & Vardell, E. (2015). Differences in the use of search assistance for tasks of varying complexity. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 23–32). ACM.
- Hu, X., & Kando, N. (2017). Task complexity and difficulty in music information retrieval. *Journal of the American Society for Information Science and Technology*, 68(7), 1711–1723.
- 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.
- Kelly, D., Arguello, J., Edwards, A., & Wu, W. (2015). Development and evaluation of search tasks for IIR experiments using a cognitive complexity framework. In *Proceedings of the International Conference on the Theory of Information Retrieval (ICTIR)* (pp. 101–110). ACM.
- O'Brien, H. L. (2016). Theoretical perspectives on user engagement. In *Why engagement matters* (pp. 1–26). Springer.
- O'Brien, H. L., & McKay, J. (2018). Modeling antecedents of engagement. In *The Handbook of Communication Engagement*. Wiley.
- O'Brien, H. L., & Toms, E. G. (2008). What is user engagement? A conceptual framework for defining user engagement with technology. *Journal of the Association for Information Science and Technology*, 59(6), 938–955.
- Wu, W. C., Kelly, D., Edwards, A., & Arguello, J. (2012). Grannies, tanning beds, tattoos and NASCAR: Evaluation of search tasks with varying levels of cognitive complexity. In *Proceedings of the International Conference on Information Interaction in Context (IIIX)* (pp. 254–257). ACM.