



# How do users examine results on Web search result pages?

Yiqun LIU

Department of Computer Science & Technology

Tsinghua University



# Search Ranking and User Feedback

-  (Google Bomb)



- **Explicit feedback is of vital importance**
- **A naïve solution: click = relevance voting**
  - Relevance = CTR (click-through rate)
- **Problem: results don't have equal opportunities**



# Search Ranking and Examination

- **How to get a justified estimation of relevance?**

- *Examination Hypothesis* (Richardson et.al, 2007)

$$C_i = 1 \rightarrow R_i = 1$$

$$C_i = 1 \rightarrow E_i = 1, R_i = 1$$

- **How to estimate the probability of examination?**

- **Fixation = examination**

- *Strong Eye-mind Hypothesis*: there is no appreciable lag between what is fixated and what is processed.  
(Just & Carpenter, 1980)



# Estimating Examination with Eye-tracking

- Eye-tracking devices



# Existing studies based on Eye-tracking

- **Users examine results with position biases**
  - Top results receive more fixations (Joachims et.al, 2005)
- **Users examine results with sequential orders**
  - However, over 50 percent of sessions still contain revisiting behaviors (Lorigo et.al, 2005)

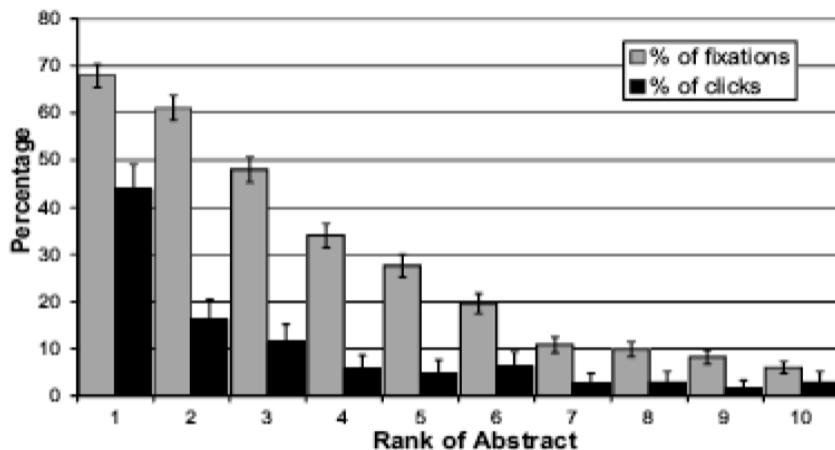


Figure 1: Percentage of times an abstract was viewed/clicked depending on the rank of the result.

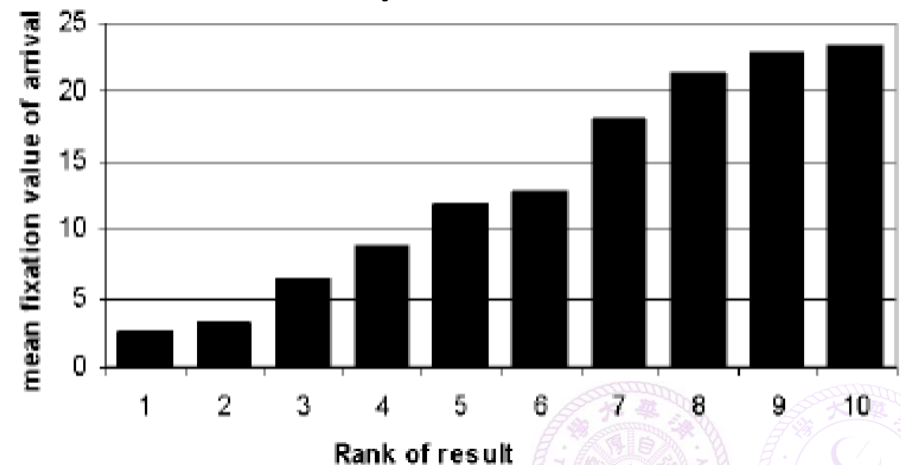
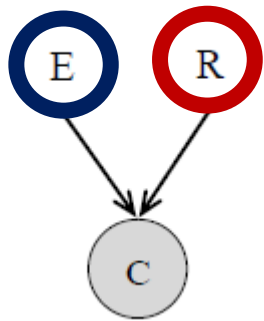


Figure 2: Mean time of arrival (in number of previous fixations) depending on the rank of the result.

# Examination beyond Eye-tracking

- Lessons learned from *Examination Hypothesis*



Relevance has to be perceived by users

Examination has to involve the comprehension of results

- Problems with *Strong Eye-mind Hypothesis*

- While the duration of the gaze is closely related to the duration of cognitive processes, the two durations are not necessarily identical. (Just & Carpenter, 1980)

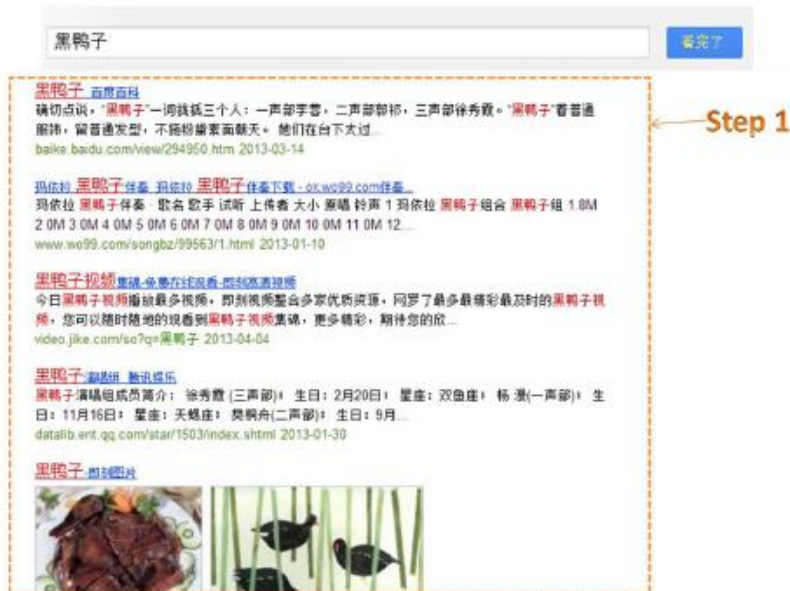
- Do fixations really mean Examination?



# Collecting Explicit Feedback on Examination

## • An Experimental Search Engine System

- Data Collected: click-through, mouse movement, eye movement, explicit feedback on examination.
- 37 participants, 25 queries (INF:TRAN:NAV = 2:2:1)



# Findings in User Feedbacks

## 1. Fixation doesn't necessarily mean examination while examination always requires fixation

	Fixation=0	Fixation=1
Examine=0	31.61%	28.81%
Examine=1	5.49%	34.09%

Why don't you annotate the fixed results as examined?

Proportion	Answers
48%	Take a glance at the result without thinking about it.
26%	Take a glance at the result and feel unattractive to read it
16%	Feel that the result is not relevant.
10%	Cannot tell clear reason.

An example of fixed results not examined



Examined  
Clicked  
220 ms fixated

Examined  
Not Clicked  
4400 ms fixated

Not Examined  
Not Clicked  
530 ms fixated

Not Examined  
Not Clicked  
380 ms fixated

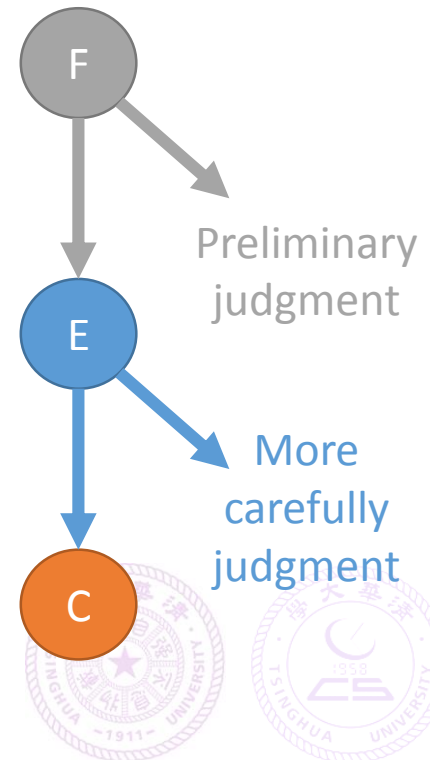
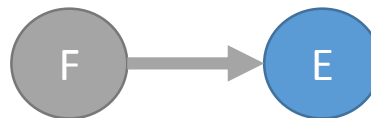
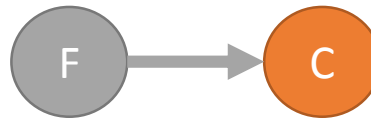
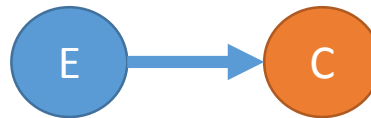
# Findings in User Feedbacks

## 2. Examination doesn't necessarily lead to click while click always requires examination

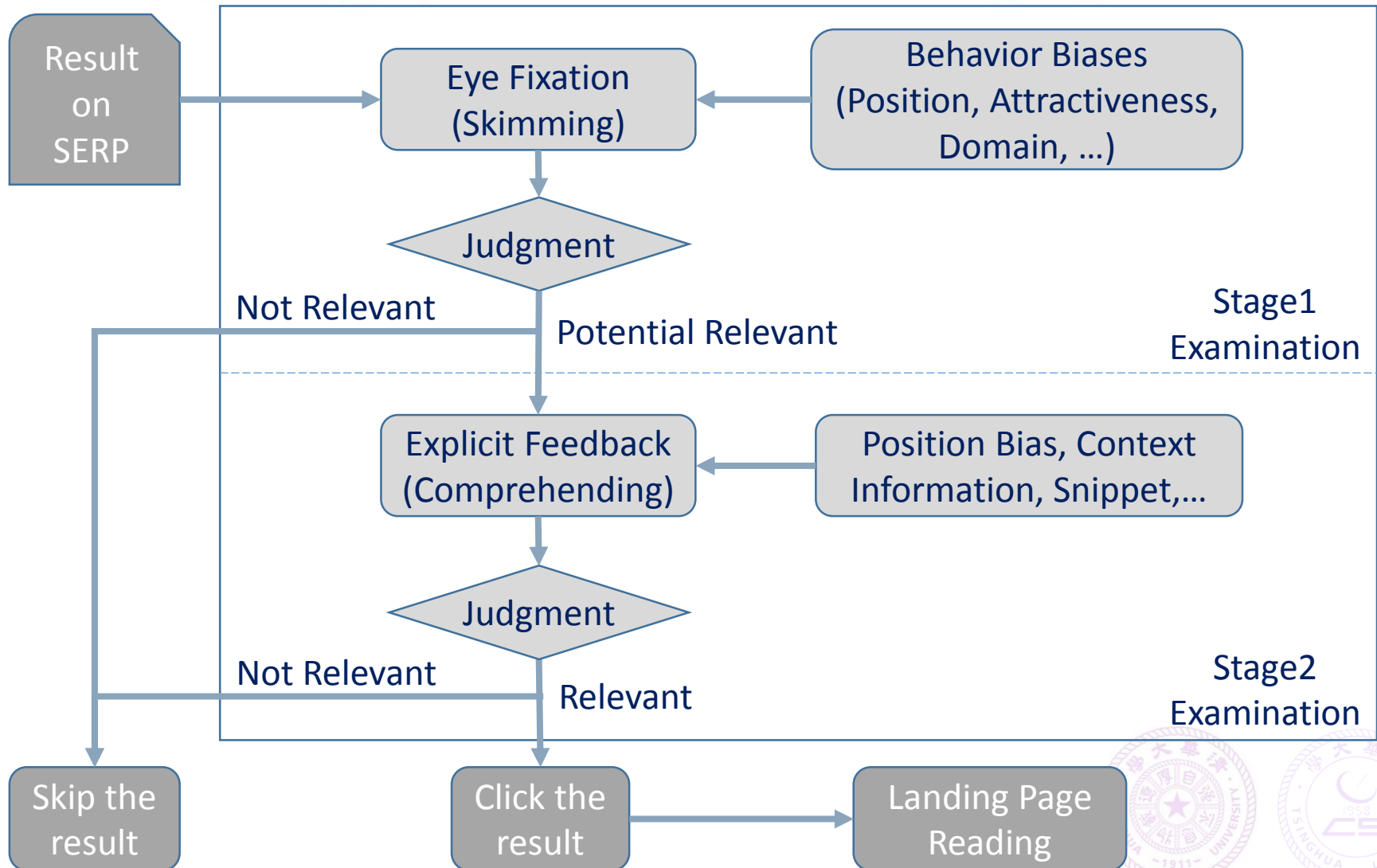
	Examine=0	Examine=1
Click=0	59.24%	17.57%
Click=1	1.18%	22.01%

	Fixation=0	Fixation=1
Click=0	34.96%	41.85%
Click=1	2.15%	21.04%

	Fixation=0	Fixation=1
Examine=0	31.61%	28.81%
Examine=1	5.49%	34.09%



# A Two-Stage Examination Model



# Behavior Biases in Two-Stage Model

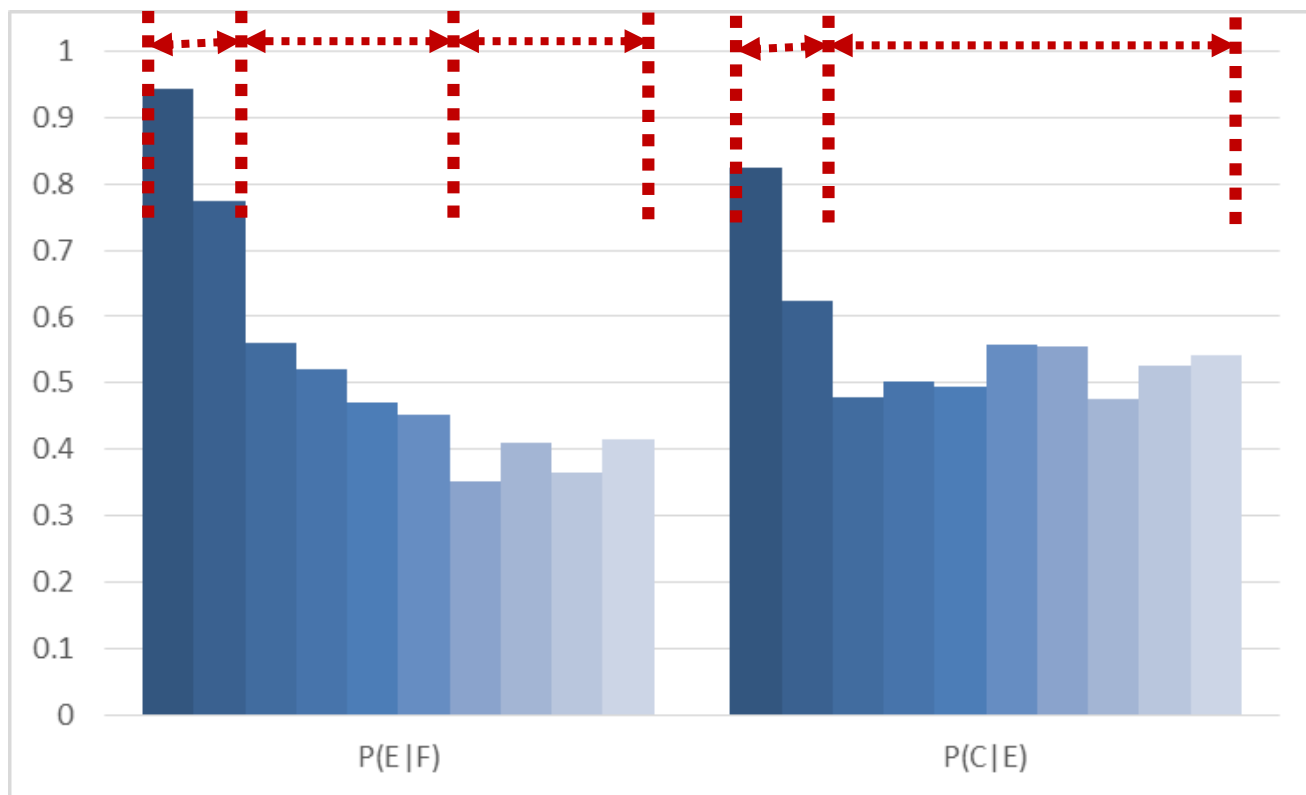
- **Behavior biases in Web search environment**

- ***Position bias***: Higher-ranked results receive more user attention (Craswell et al. 2008)
- ***Attractiveness bias***: attractiveness in result titles and abstracts affects user judgment (Bar-Ilan et al. 2009), multimedia vertical results draws much user attentions (Wang et al. 2013)
- ***Trust bias***: Results from trust-worthy Web domains are preferred by users (leong et al. 2012)



# Position Bias in Two Stages of Examination

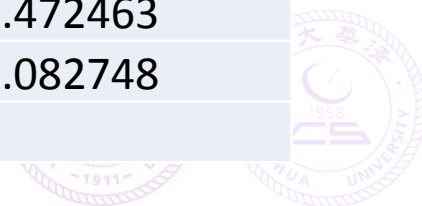
- User judgments (for relevant results) in two stages are both affected by positions



# Attractiveness Bias in Two Stages of Examination

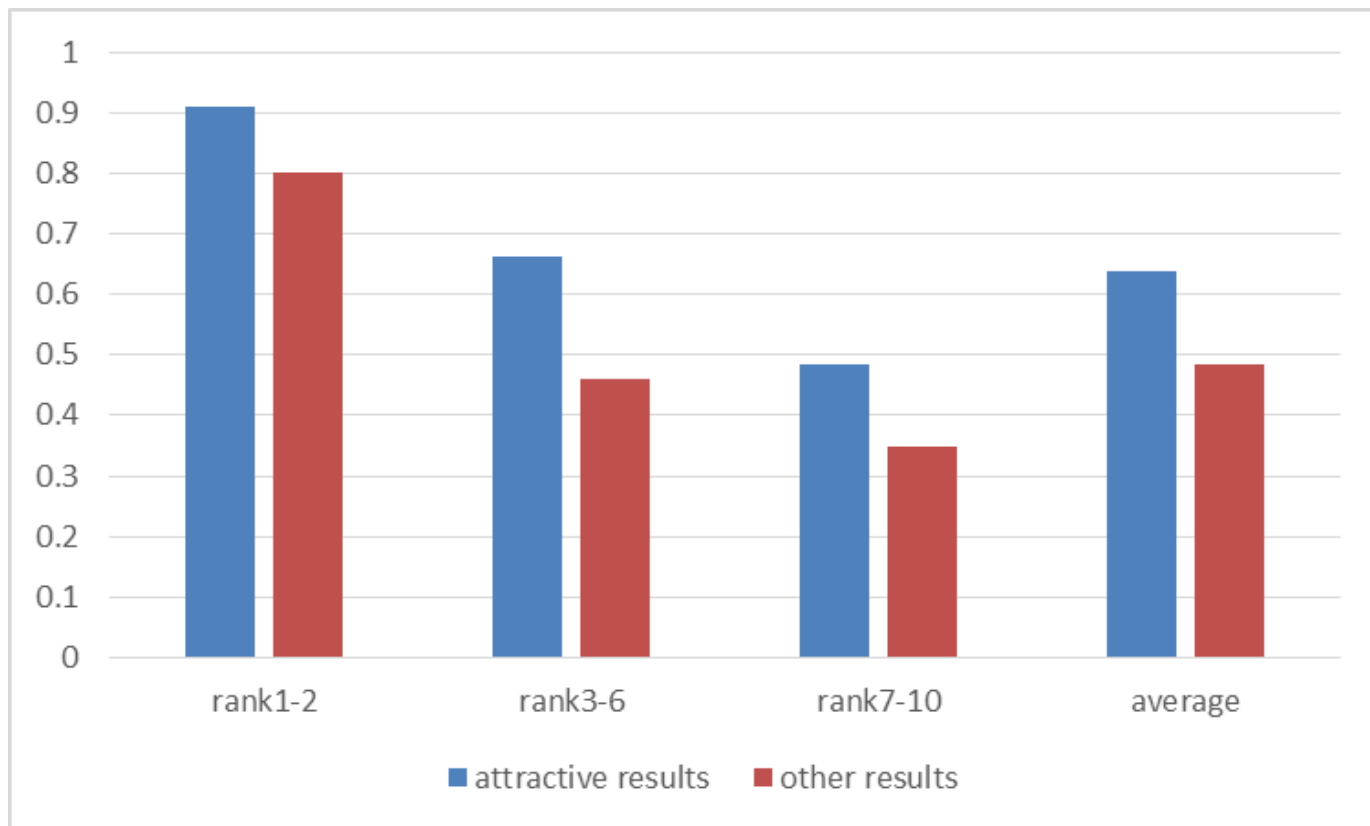
- **Attractive results draws significantly more attention in Stage 1 while doesn't affect the judgment in Stage 2.**
- Attractive results: top 1/2/3/5 results with the highest title/abstract exact match on a SERP

		Attractive results	Other results
P(E F)	Average	0.637301	0.484615
	Variance	0.058769	0.066037
	p-value	0.005788	
P(C E)	Average	0.57775	0.472463
	Variance	0.122599	0.082748
	p-value	0.158477	



# Attractiveness Bias v.s. Position Bias

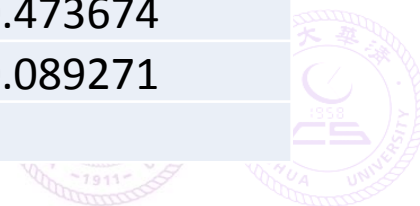
- **Attractiveness bias happens in all result positions for judgments in Stage 1.**



# Trust Bias in Two Stages of Examination

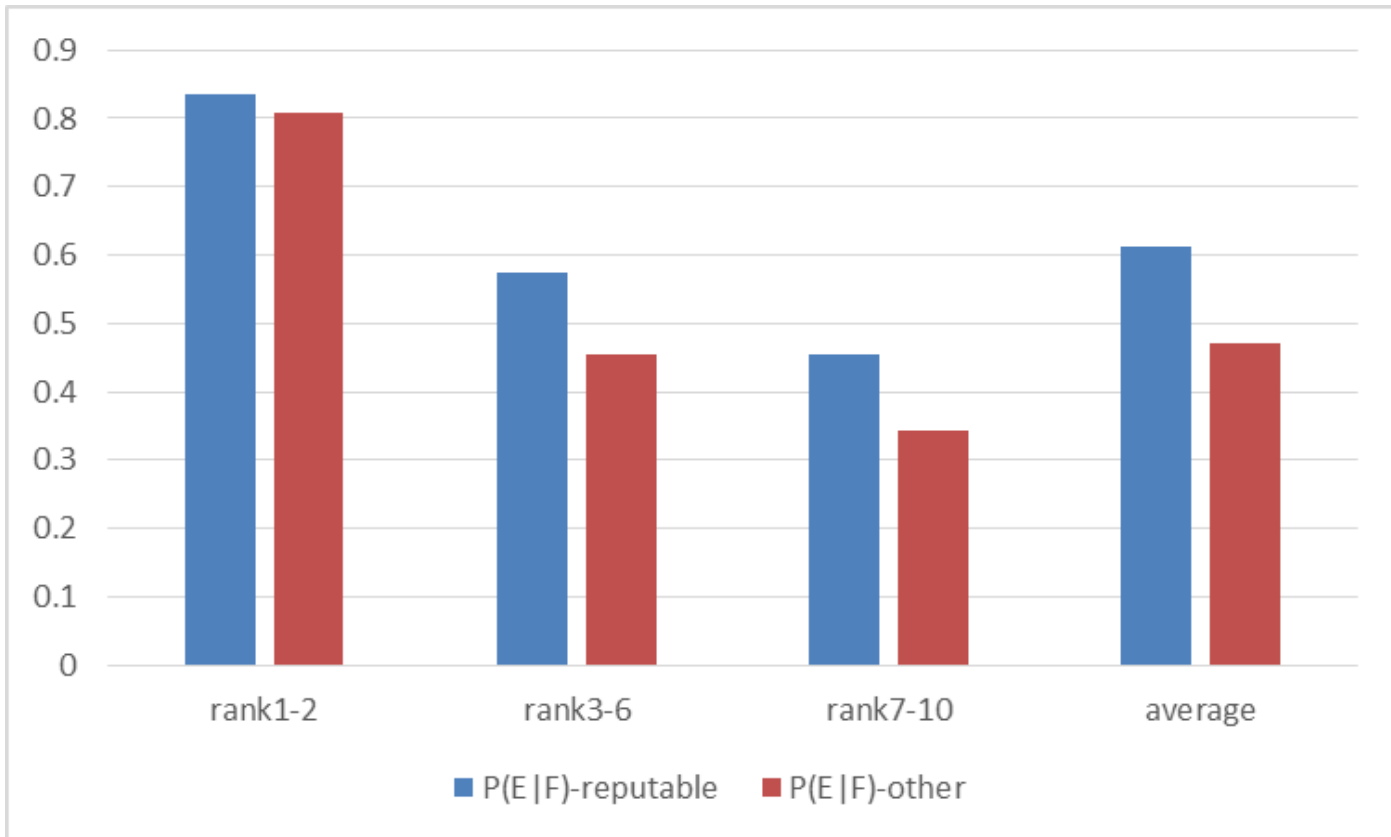
- **Reputable results draws significantly more attention in Stage 1 while doesn't affect the judgment in Stage 2.**
- Reputable results: results from Alexa.com top 100 popular sites in China

		Attractive results	Other results
P(E F)	Average	0.613371	0.519443
	Variance	0.065817	0.079853
	p-value	0.000656	
P(C E)	Average	0.470799	0.473674
	Variance	0.063693	0.089271
	p-value	0.311937	



# Trust Bias v.s. Position Bias

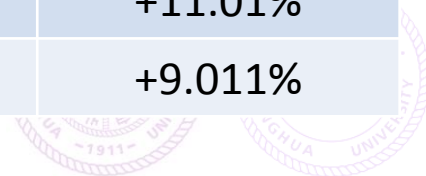
- Trust bias happens in relatively lower result positions for judgments in Stage 1.



# Effectiveness of Judgments in Two Stages

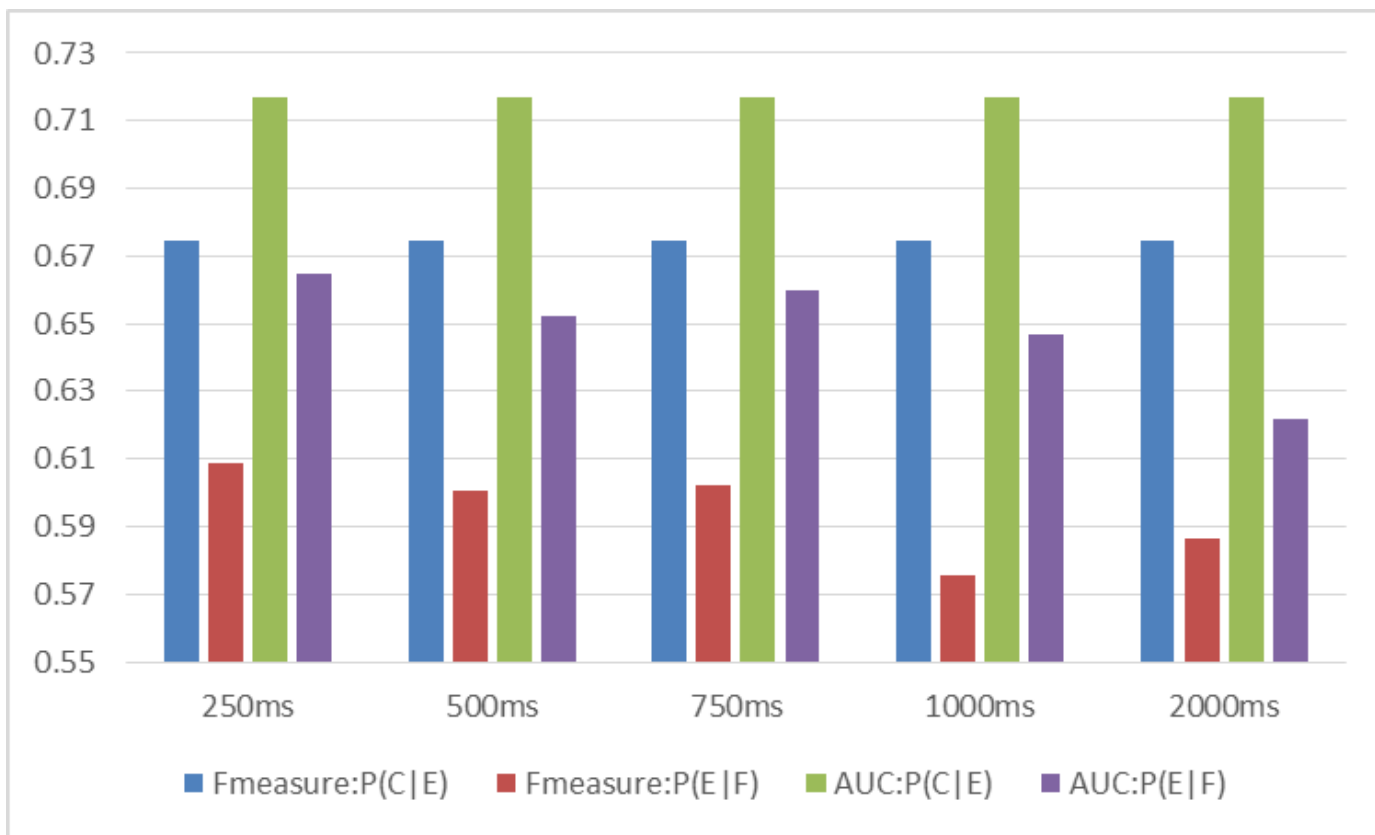
- **User examines more results in Stage 1, but the effectiveness of judgments in Stage 2 is higher**
  - Relevance judgment in Stage 1: entering Stage 2
  - Relevance judgment in Stage 2: result clicking

	Stage 1	Stage 2	Comparison
Number of examined results	5598/8900	3034/5598	-45.80%
Number of results judged as relevant	3034/5598	1873/3034	-38.27%
Precision	0.5968	0.6738	+11.43%
Recall	0.6040	0.6755	+10.58%
F-measure	0.6004	0.6747	+11.01%
AUC/ROC	0.6523	0.7169	+9.011%



# Effectiveness of Judgments in Two Stages

- Effectiveness comparison results do not change with fixation threshold settings



# Discussion



# Take-home Messages

- **1. Users examine results with a two-stage model**
  - Stage1: from skimming to comprehending, judging whether he/she should carefully read the result
  - Stage2: from comprehending to clicking, judging whether he/she should click the result and obtain information from the landing page
- **2. Behavior biases happen in different stages**
- **3. User examines more results in Stage 1, but the effectiveness of judgments in Stage 2 is higher**



# Thank you



请各位批评指正！

**Welcome to visit my homepage**

<http://www.thuir.cn/group/~YQLiu/>

