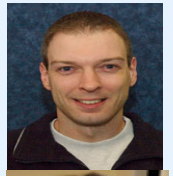
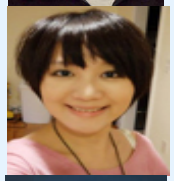
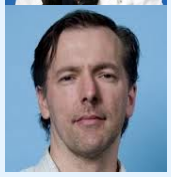
The Effects of Threshold Priming and Need for Cognition on Relevance Assessments

Diane Kelly, Associate Professor
School of Information & Library Science
University of North Carolina at Chapel Hill







In collaboration with:

- ✧ Falk Scholer, RMIT University, Australia
- ✧ Wan-ching Wu, University of North Carolina at Chapel Hill
- ✧ Hanseul Lee, University of North Carolina at Chapel Hill
- ✧ William Webber, University of Maryland

Scholer, F., Kelly, D., Wu, W.-C., Lee, H. & Webber, W. (2013). The effects of threshold priming and Need for Cognition on relevance calibration and assessment. *Proceedings of SIGIR '13*, 623-632.

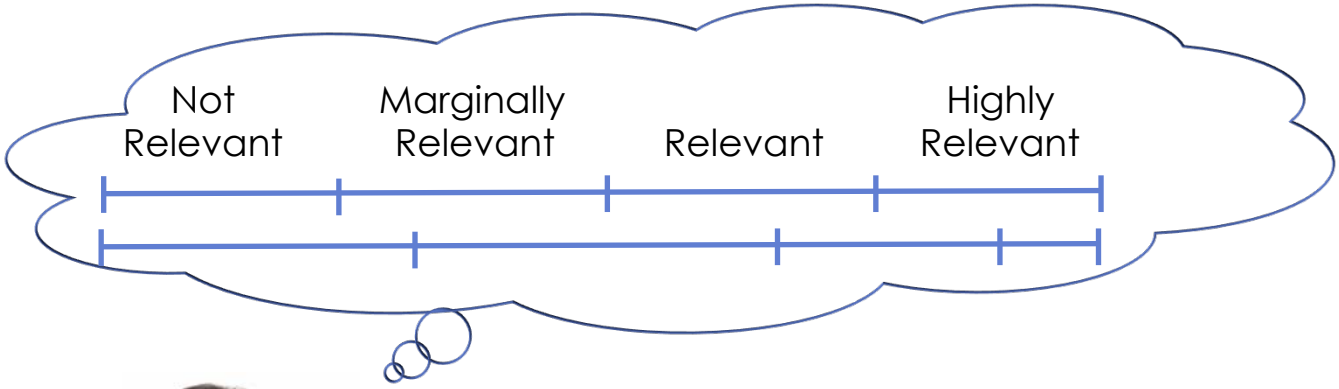
Problem

- Relevance is a fundamental concept in information science and in information retrieval.
- Relevance assessments are needed for a variety of tasks including the development of research infrastructure (test collections, e.g., TREC), ad-hoc evaluation, and training text classifiers.
- It is generally understood that there is great variability in the relevance assessments that people make and that assessments are not independent.
- Research has shown this variability might be okay in situations where relative comparisons are made but not in cases where absolute evaluation is needed.

Relevance

- Saracevic's Five Types of Relevance
  - Algorithmic
  - Topical
  - Cognitive
  - Situational
  - Affective
- Harter's psychological relevance, ". . . relevance judgments are a function of one's mental state at the time a reference is read. They are not fixed; they are dynamic." (p. 612)

Threshold Priming



Need for Cognition

- A stable individual difference in people's tendency to engage in and enjoy effortful cognitive activity (Cohen, Stotland and Wolfe,1955; Cacioppo & Petty, 1982)
- Empirical studies have found that individuals with high Need for Cognition are more motivated to process information, pay more attention to argument quality, perform better on cognitive tasks and react more positively to complex rules

Research Questions

- Does threshold priming affect how participants make relevance assessments?
- Does Need for Cognition affect how participants make relevance assessments?
- Does threshold priming impact participants' confidence in their relevance assessments?
- What do participants find challenging about deciding which relevance levels to associate with documents?

Method

- Laboratory experiment with one between-subjects independent variable that had three levels: high, medium and low.
- Each participant was given one search topic and asked to evaluate the relevance of 48 documents.
- The independent variable, or treatment, was 'administered' by manipulating the relevance levels of the first 20 documents participants encountered in the list of 48.

Experimental Treatments

Prologue				Epilogue					
1	X	11	NR	21	R-MR	31	HR	41	R-MR
2	X	12	X	22	R-MR	32	R-MR	42	NR
3	NR	13	NR	23	NR	33	R-MR	43	R-MR
4	NR	14	X	24	R-MR	34	R-MR	44	NR
5	X	15	NR	25	R-MR	35	HR	45	HR
6	NR	16	X	26	HR	36	R-MR	46	R-MR, 21
7	X	17	X	27	NR	37	HR	47	R-MR, 22
8	NR	18	NR	28	HR	38	R-MR	48	R-MR, 24
9	X	19	NR	29	NR	39	NR		
10	NR	20	X	30	NR	40	HR		

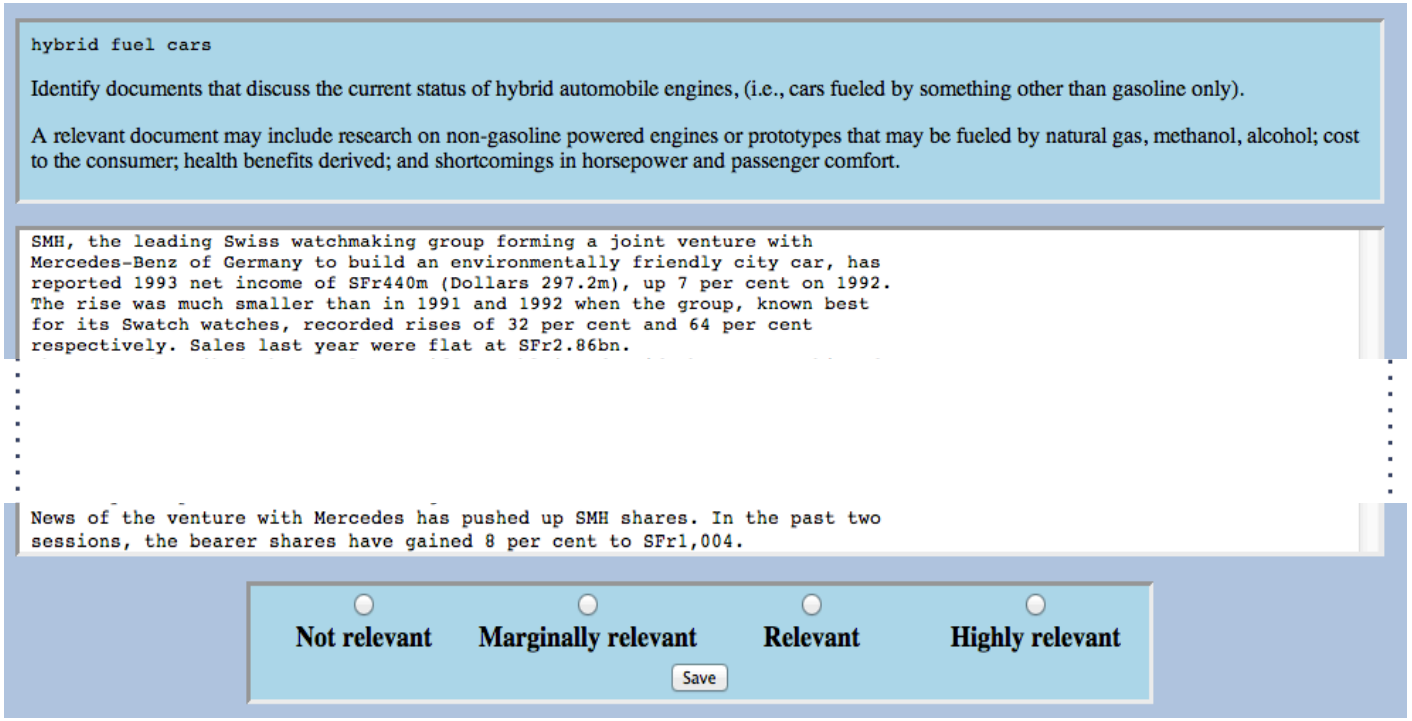
- *X* = documents that differed in the *Prologue* according to treatment.
- *NR* = non-relevant document
- *MR-R* = relevant or marginally relevant document
- *HR* = highly relevant

- Treatments (Between subjects)
  - High Treatment: X➔HR
  - Medium Treatment: X➔MR-R
  - Low Treatment: X➔NR
- Documents in *Epilogue* were identical.
- Documents in positions 46-48 were duplicates of those in positions 21, 22 & 24.

Relevance Categories

- **Highly Relevant (3)**: The document discusses the themes of the topic exhaustively. In case of a multi-faceted topic, all or most sub-themes or viewpoints are covered. Typical extent: several text paragraphs, at least 4 sentences or facts.
- **Relevant (2)**: The document contains more information than the topic description but the presentation is not exhaustive. In case of a multi-faceted topic, only some of the sub-themes or viewpoints are covered. Typical extent: one text paragraph, 2-3 sentences or facts.
- **Marginally relevant (1)**: The document only points to the topic. It does not contain more or other information than the topic description. Typical extent: one sentence or fact.
- **Not relevant (0)**: The document does not contain any information about the topic.

Assessment Interface



Search Topics

	Hybrid Fuel Cars	Sick Building Syndrome	Drugs, Golden Triangle
Frequency of past search	Never	Never	Never
Prior Knowledge	A little	A little	Nothing
Interest	Slightly	Slightly	Somewhat
Relevance to life	Moderately	Slightly	Not at all

Exit Questionnaire

- What, if anything, was challenging about deciding which relevance levels to associate with each document?
- How confident are you in the relevance judgments you made (not at all confident, slightly confident, somewhat confident, moderately confident, very confident)?
- Need for Cognition Scale (Cacioppo & Petty, 1982)
  - 18-items; 5-point scale
  - "The notion of thinking abstractly is appealing to me."
  - "I like to have the responsibility of handling a situation that requires a lot of thinking."

Participants

- 82 participants from university community
- Ages ranged from 18 to 25 (mean = 23.70)
- 85% female
- 85% students
- 87% native English speakers

Results

Does threshold priming affect how participants make relevance assessments?

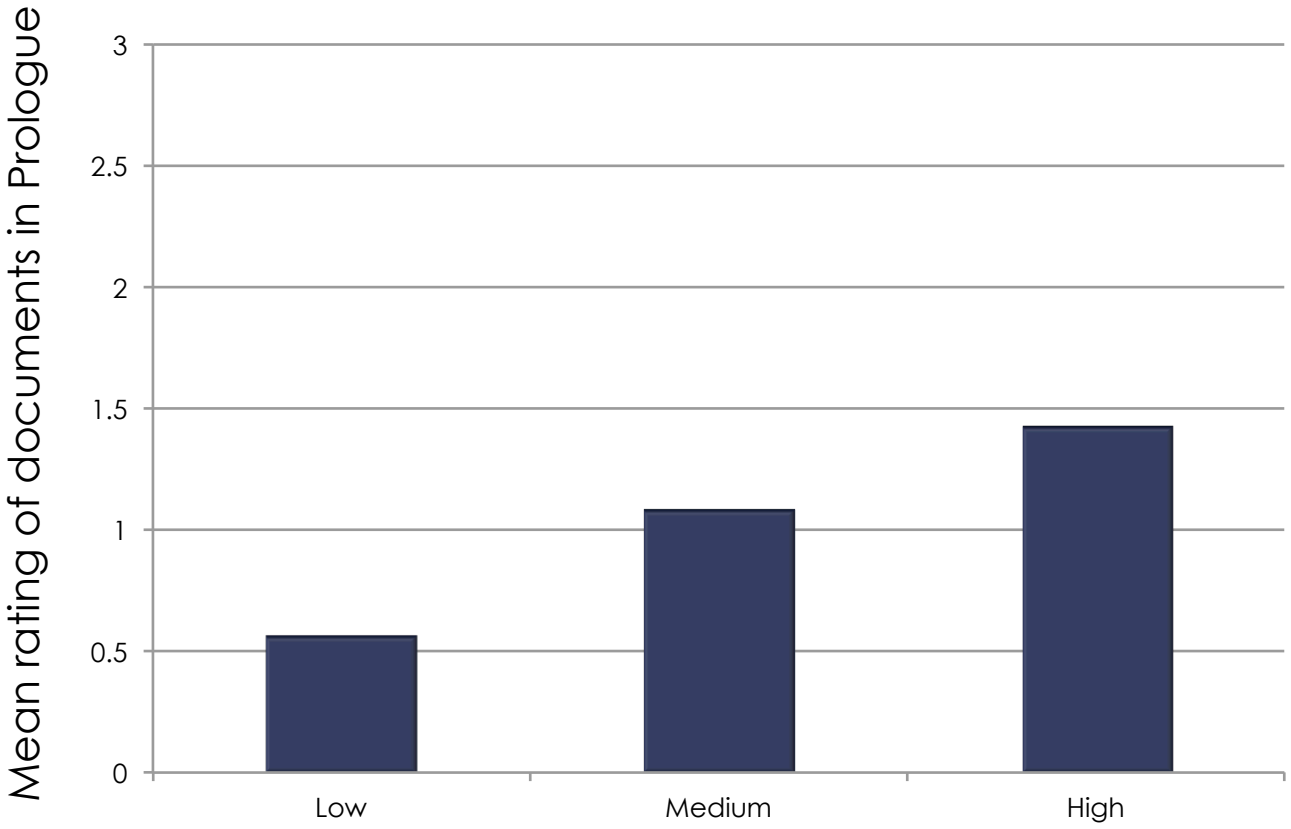
Experimental Treatments

Prologue				Epilogue					
1	X	11	NR	21	R-MR	31	HR	41	R-MR
2	X	12	X	22	R-MR	32	R-MR	42	NR
3	NR	13	NR	23	NR	33	R-MR	43	R-MR
4	NR	14	X	24	R-MR	34	R-MR	44	NR
5	X	15	NR	25	R-MR	35	HR	45	HR
6	NR	16	X	26	HR	36	R-MR	46	R-MR, 21
7	X	17	X	27	NR	37	HR	47	R-MR, 22
8	NR	18	NR	28	HR	38	R-MR	48	R-MR, 24
9	X	19	NR	29	NR	39	NR		
10	NR	20	X	30	NR	40	HR		

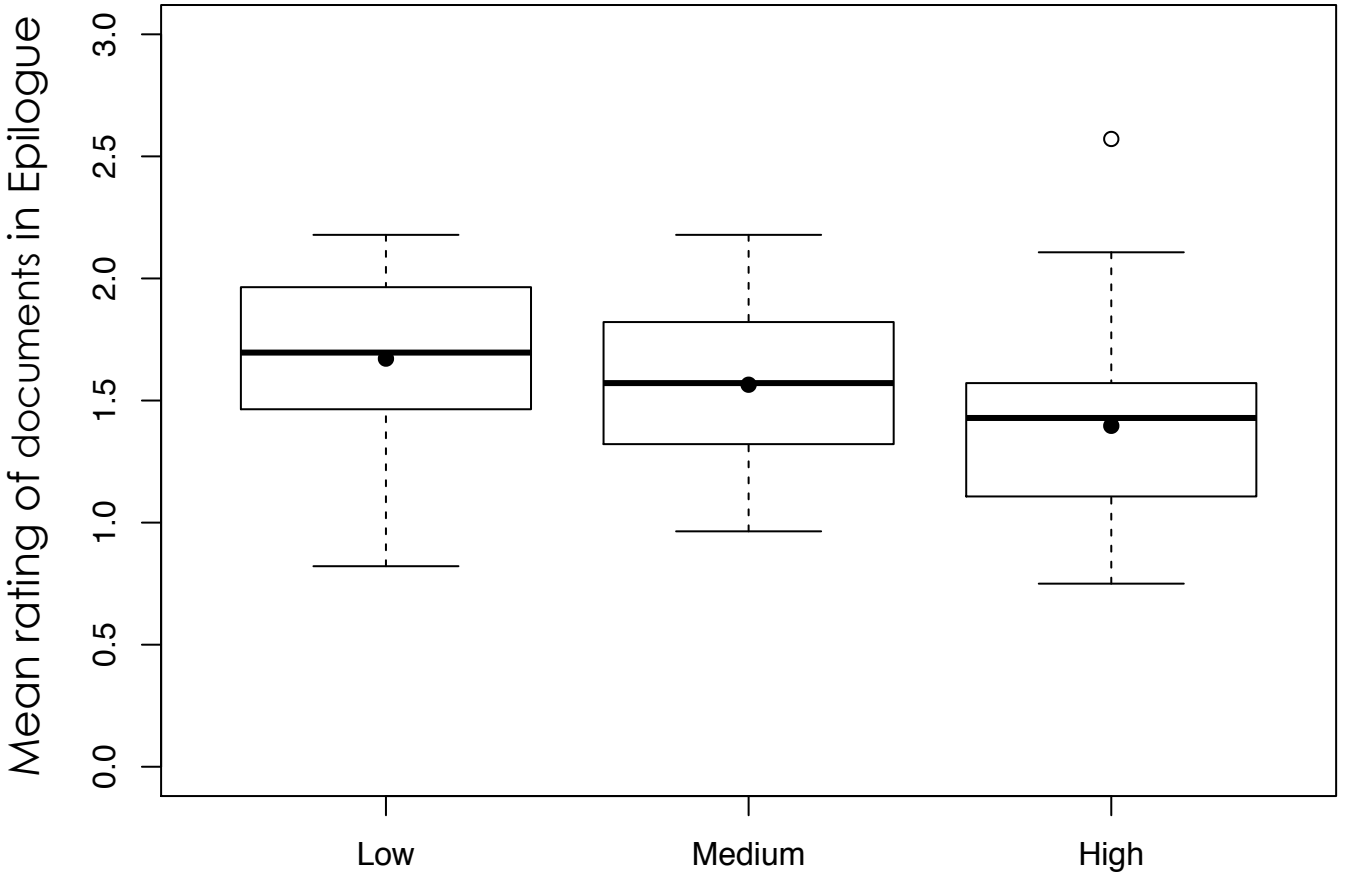
- Treatments (Between subjects)
  - High Treatment: X➔HR
  - Medium Treatment: X➔MR-R
  - Low Treatment: X➔NR

- NR documents in Prologue identical.
- All documents in *Epilogue* were identical.

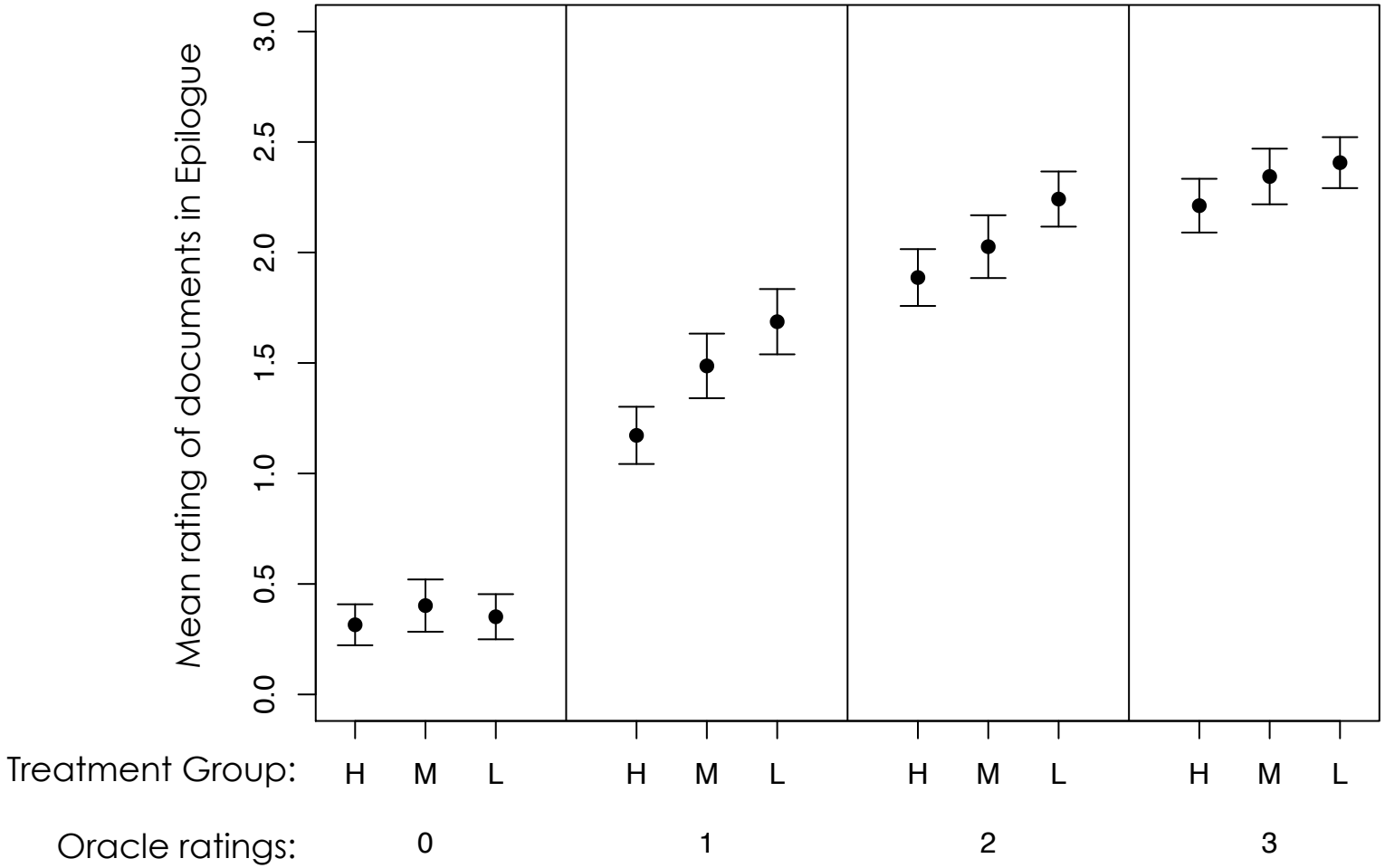
Treatment Check



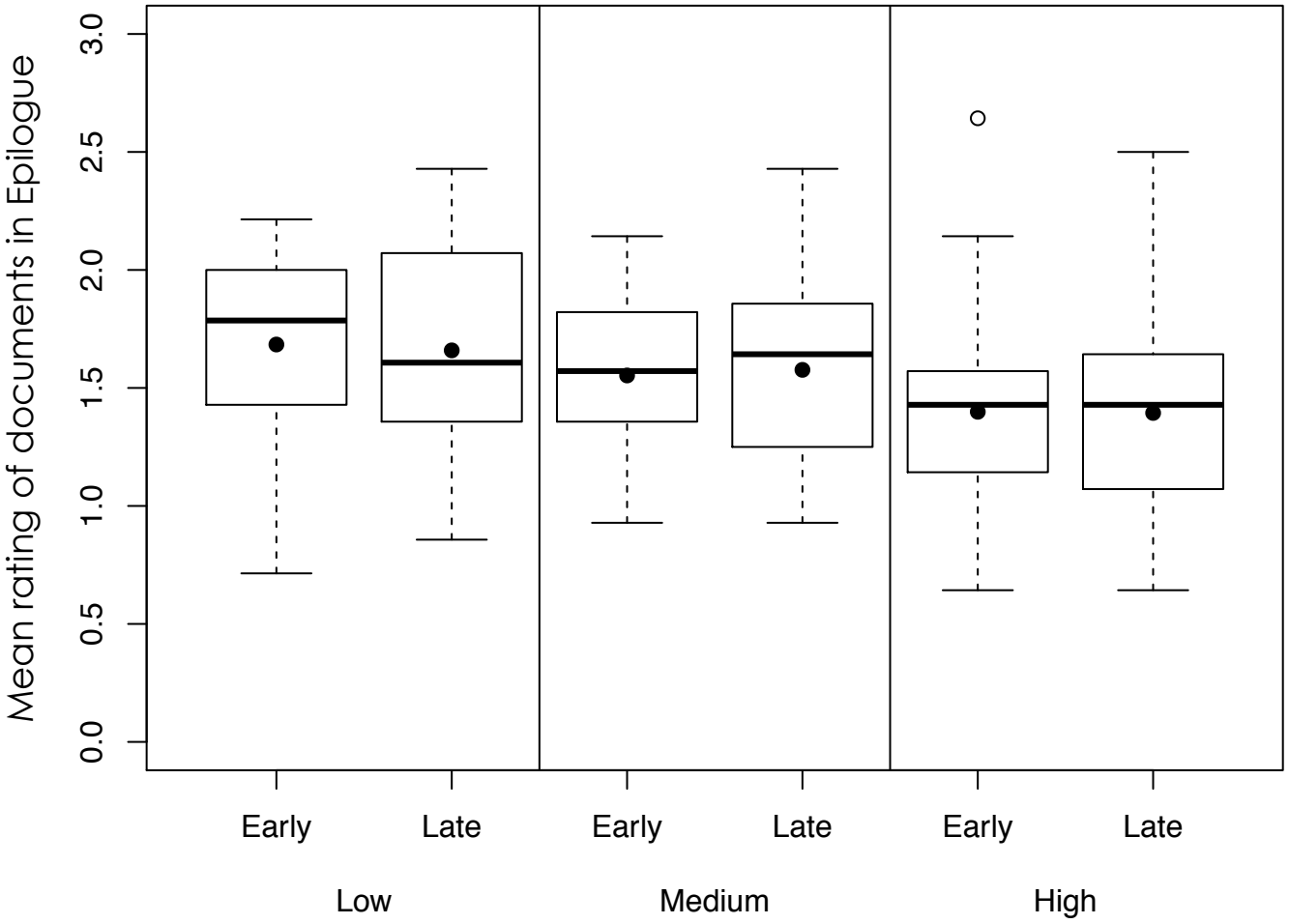
Treatment Effect



Treatment Effect



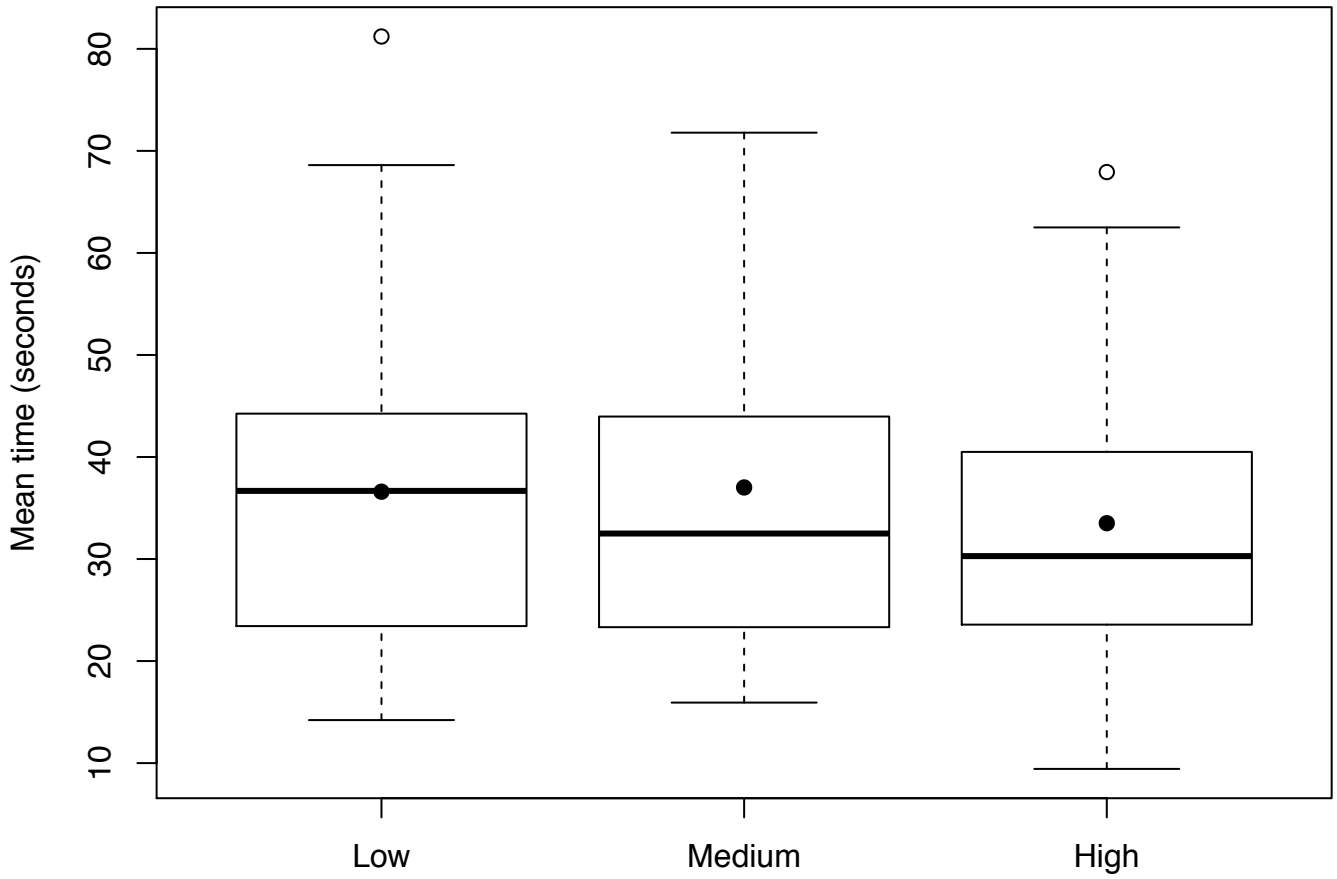
Epilogue: Early and Late Judgments



Assessments of Non-relevant Documents

	Low	Medium	High
Prologue	0.65	0.48	0.33
Epilogue	0.35	0.42	0.32
Oracle	0	0	0

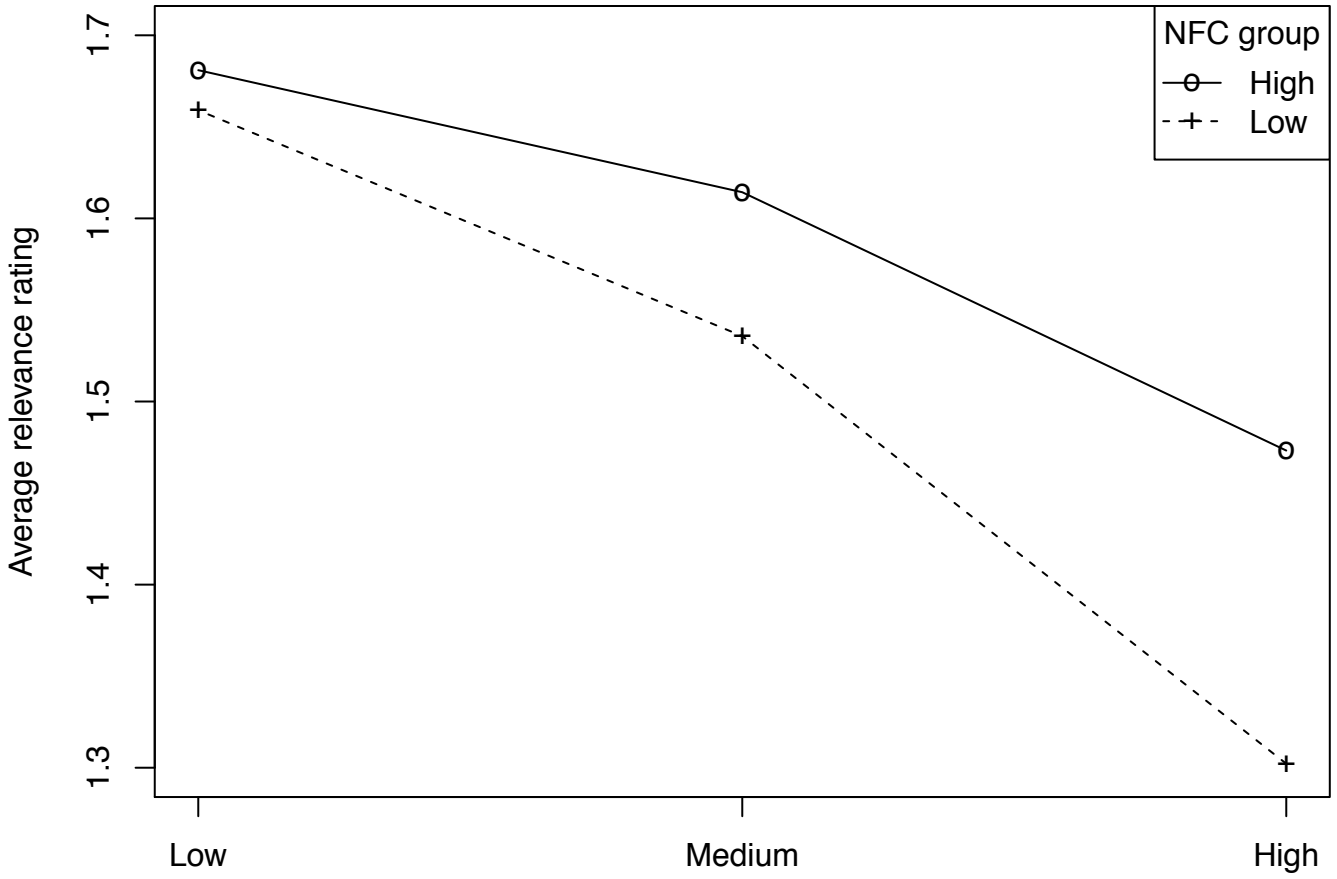
Epilogue: Time Taken to Judge



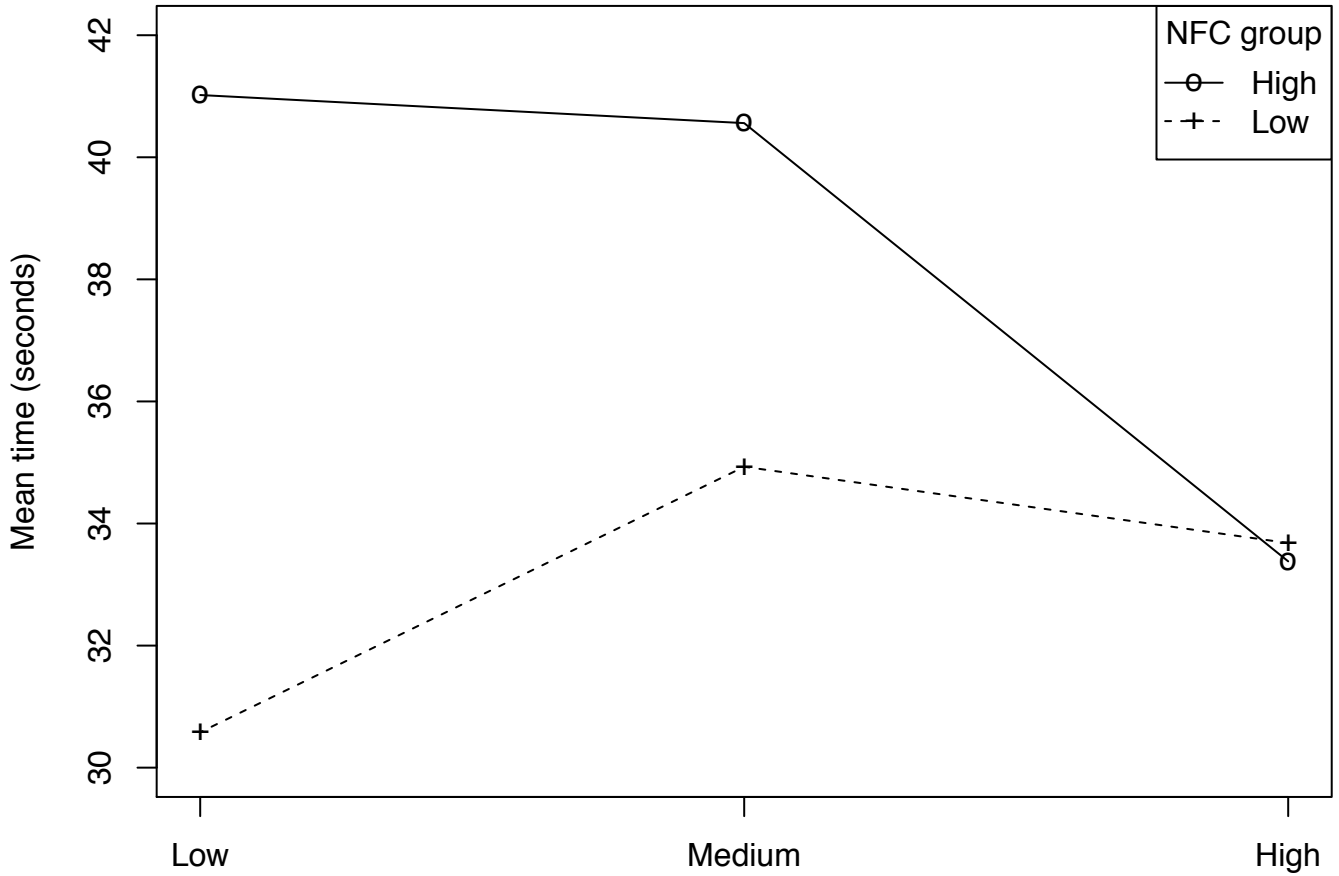
Results

Does Need for Cognition affect how participants make relevance assessments?

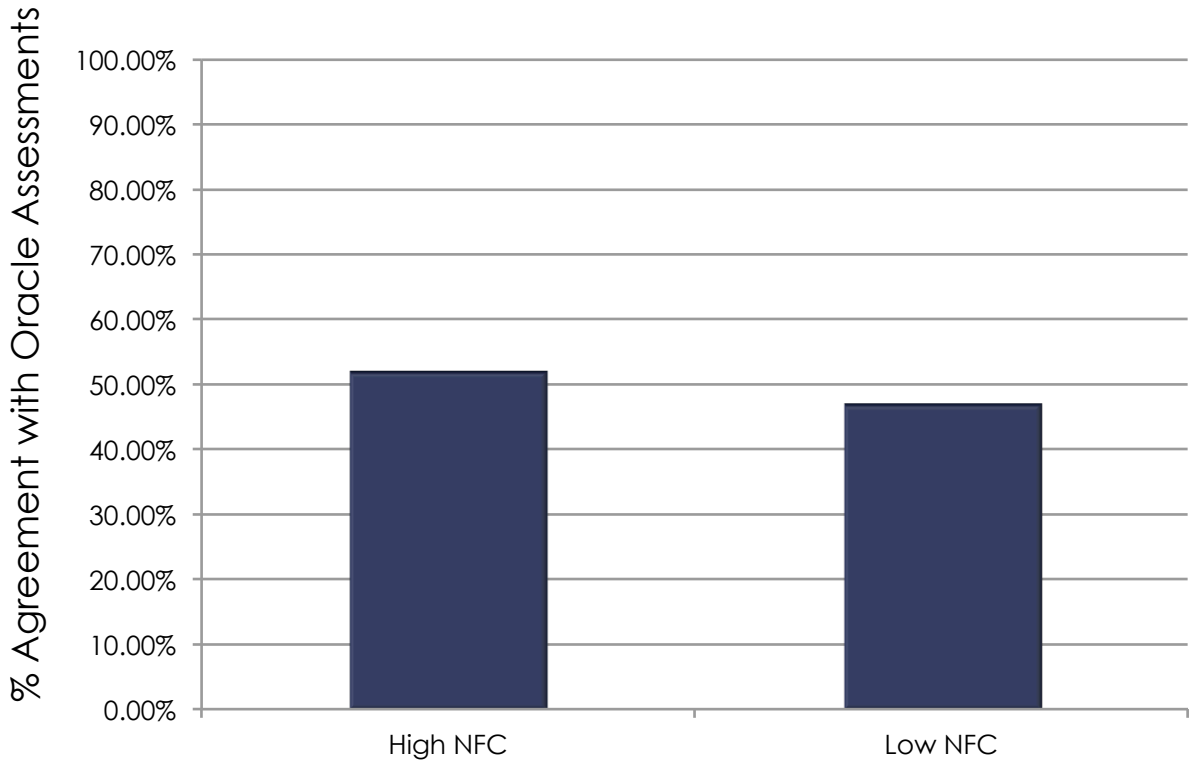
Need for Cognition and Judgments



Need for Cognition and Time



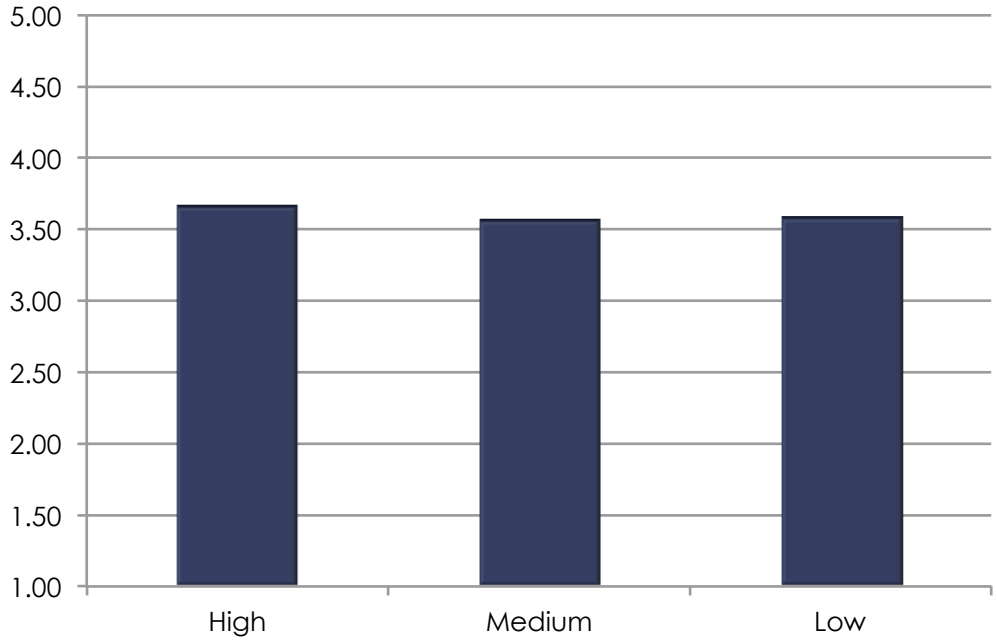
Need for Cognition and Accuracy



Results

Does threshold priming impact participants' confidence in their relevance assessments?

Confidence in Judgments



5 = very confident
3 = somewhat confident
1 = not at all confident

Results

What did participants find challenging about deciding which relevance levels to associate with documents?

Challenges

- Determining the extent to which a topic was represented in a document.
- Difficulty separating *topical* relevance from other types of relevance. Participants made comments about:
  - Cognitive relevance
  - Situational relevance
  - Affective relevance

Summary and Implications

- Participants in the low treatment group assigned significantly higher mean relevance scores to documents in the epilogue than participants in the high treatment group.
  - Biggest differences in scores were assigned to documents encountered early in the epilogue.
- Although high Need for Cognition participants assigned higher relevance scores to documents and spent longer time assessing documents, these differences were not significant.
- High Need for Cognition participants made more 'accurate' relevance assessments.
- Restricting one's assessments to topical relevance is difficult, if not impossible.

Extensions

- Capture people's reasons for their relevance ratings in real-time
- Examine if people's articulation of reasons for relevance ratings mitigates or enhances treatment effects
- Examine if people's reasons differ according to treatment
- Examine potential anchoring effects when people are assessing documents for a multi-faceted search topic

Thank you.

Ask questions now or email me later: **dianek@email.unc.edu**