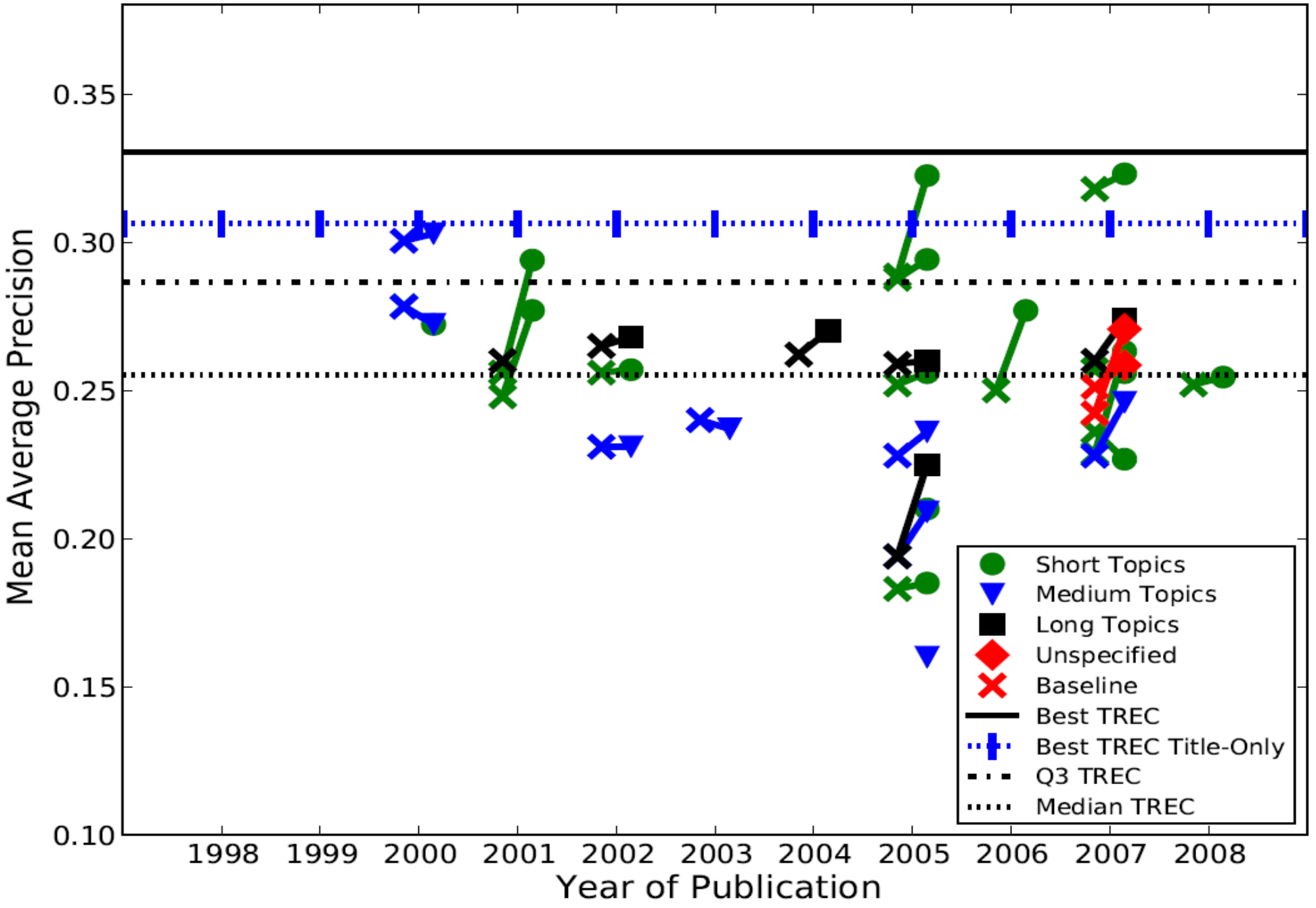
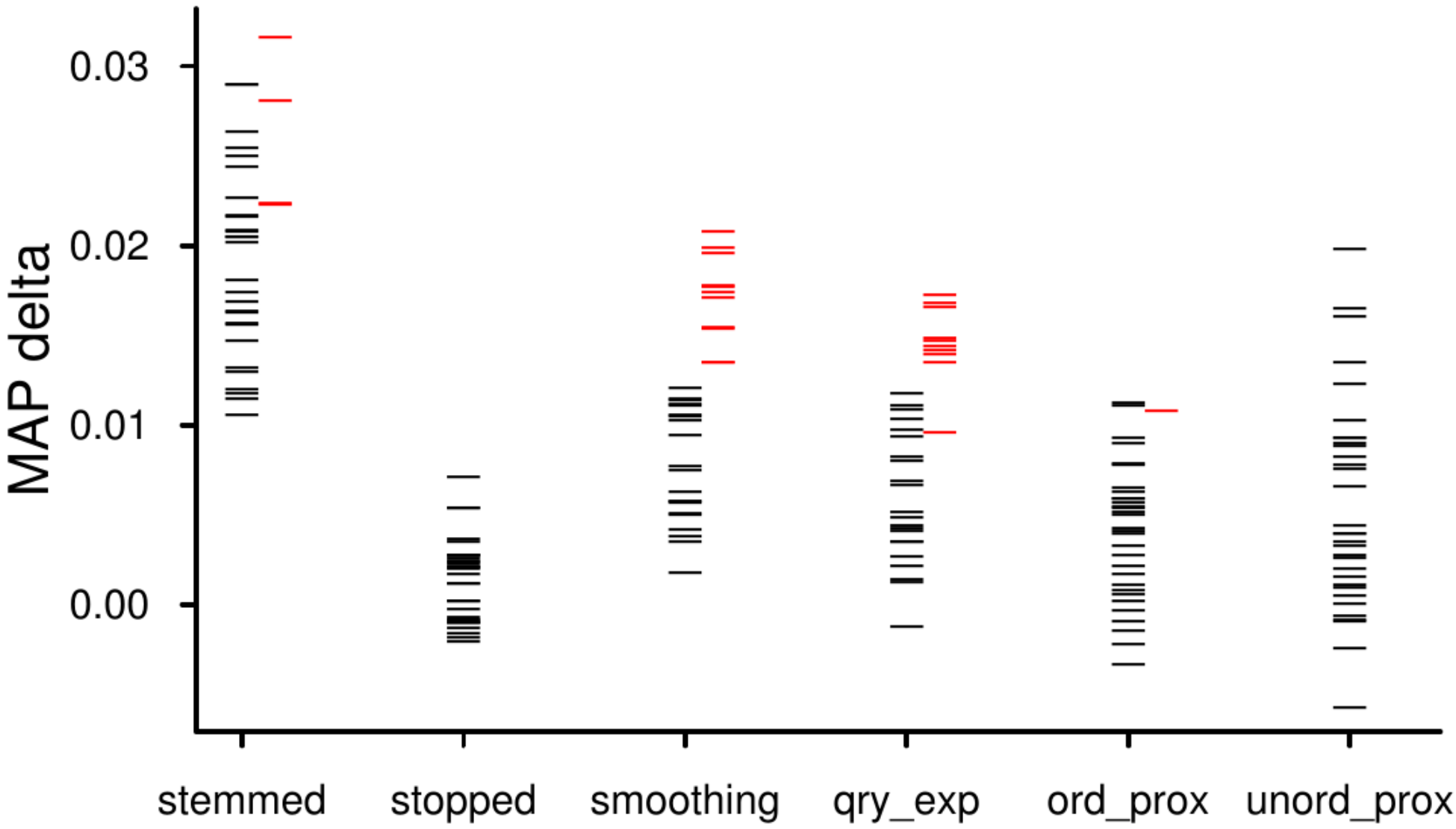
Causality, prediction and improvements that (don't) add up

Norbert Fuhr
University of Duisburg-Essen
Germany
norbert.fuhr@acm.org

Armstrong et al 2009: Improvements that don‘t add up



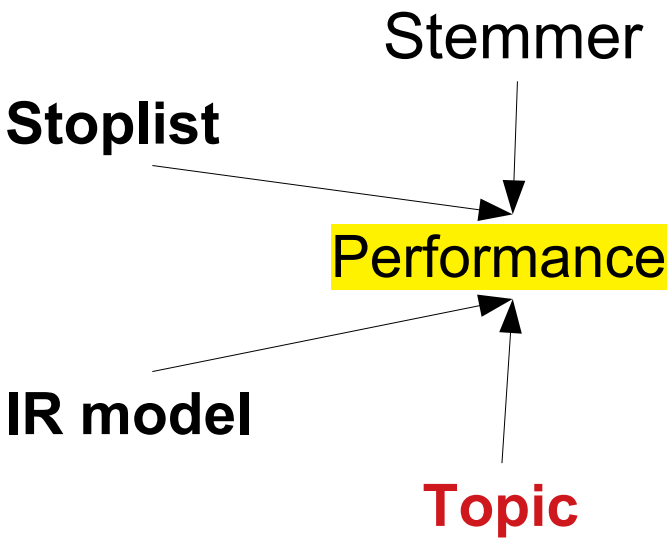
(a) TREC-8 Ad-Hoc ($n = 32$)



Significance of option combinations on TREC5

[Armstrong et al. 2009]

Ferro & Silvello 2016: A General Linear Mixed Models Approach to Study System Component Effects



- ANOVA analysis of the effect of different factors on system performance
- Biggest variance due to topics
- Remaining error still large
- → suitability of linear model?

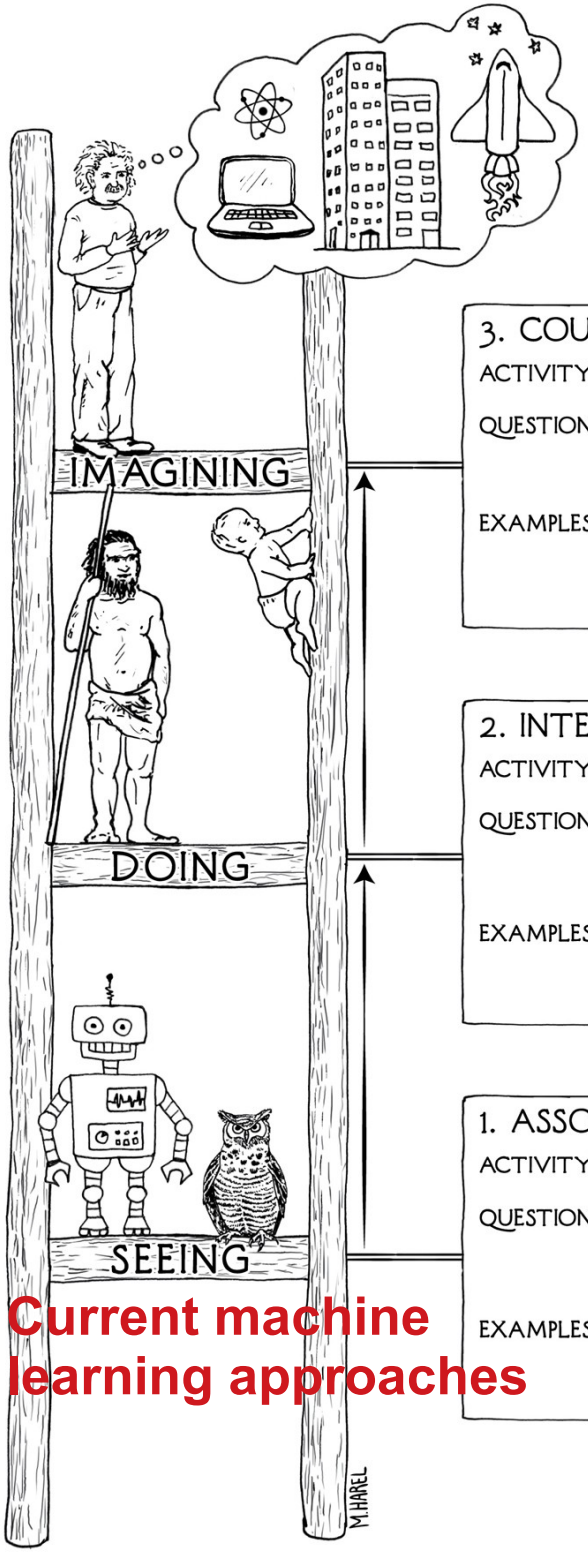
JUDEA PEARL

WINNER OF THE TURING AWARD

AND DANA MACKENZIE

THE BOOK OF WHY

THE NEW SCIENCE OF CAUSE AND EFFECT



3. COUNTERFACTUALS

ACTIVITY: Imagining, Retrospection, Understanding

QUESTIONS: *What if I had done …? Why?* (Was it X that caused Y? What if X had not occurred? What if I had acted differently?)

EXAMPLES: Was it the aspirin that stopped my headache? Would Kennedy be alive if Oswald had not killed him? What if I had not smoked for the last 2 years?

2. INTERVENTION

ACTIVITY: Doing, Intervening

QUESTIONS: *What if I do …? How?* (What would Y be if I do X? How can I make Y happen?)

EXAMPLES: If I take aspirin, will my headache be cured? What if we ban cigarettes?

1. ASSOCIATION

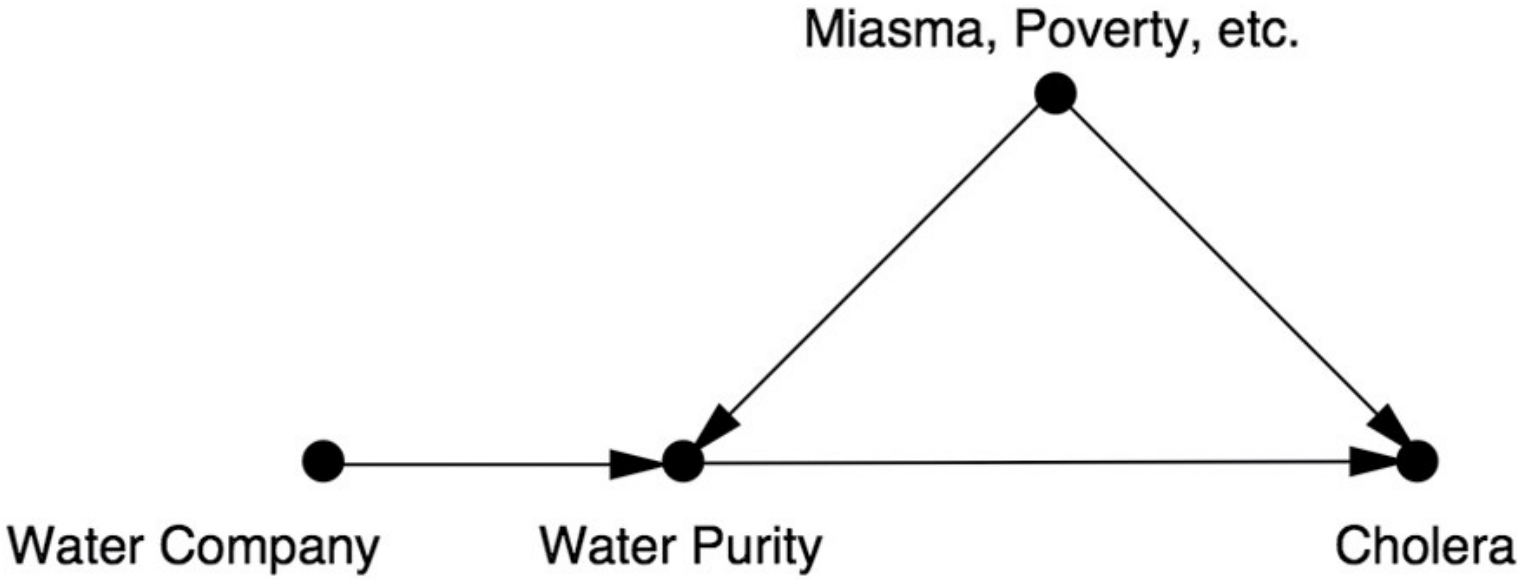
ACTIVITY: Seeing, Observing

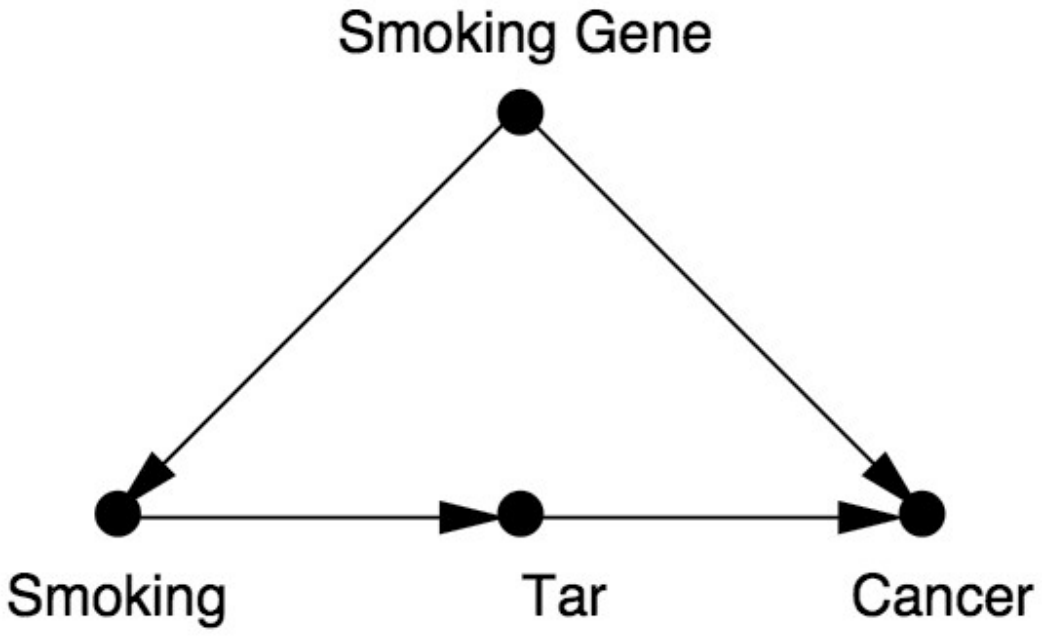
QUESTIONS: *What if I see ...?* (How are the variables related? How would seeing X change my belief in Y?)

EXAMPLES: What does a symptom tell me about a disease? What does a survey tell us about the election results?

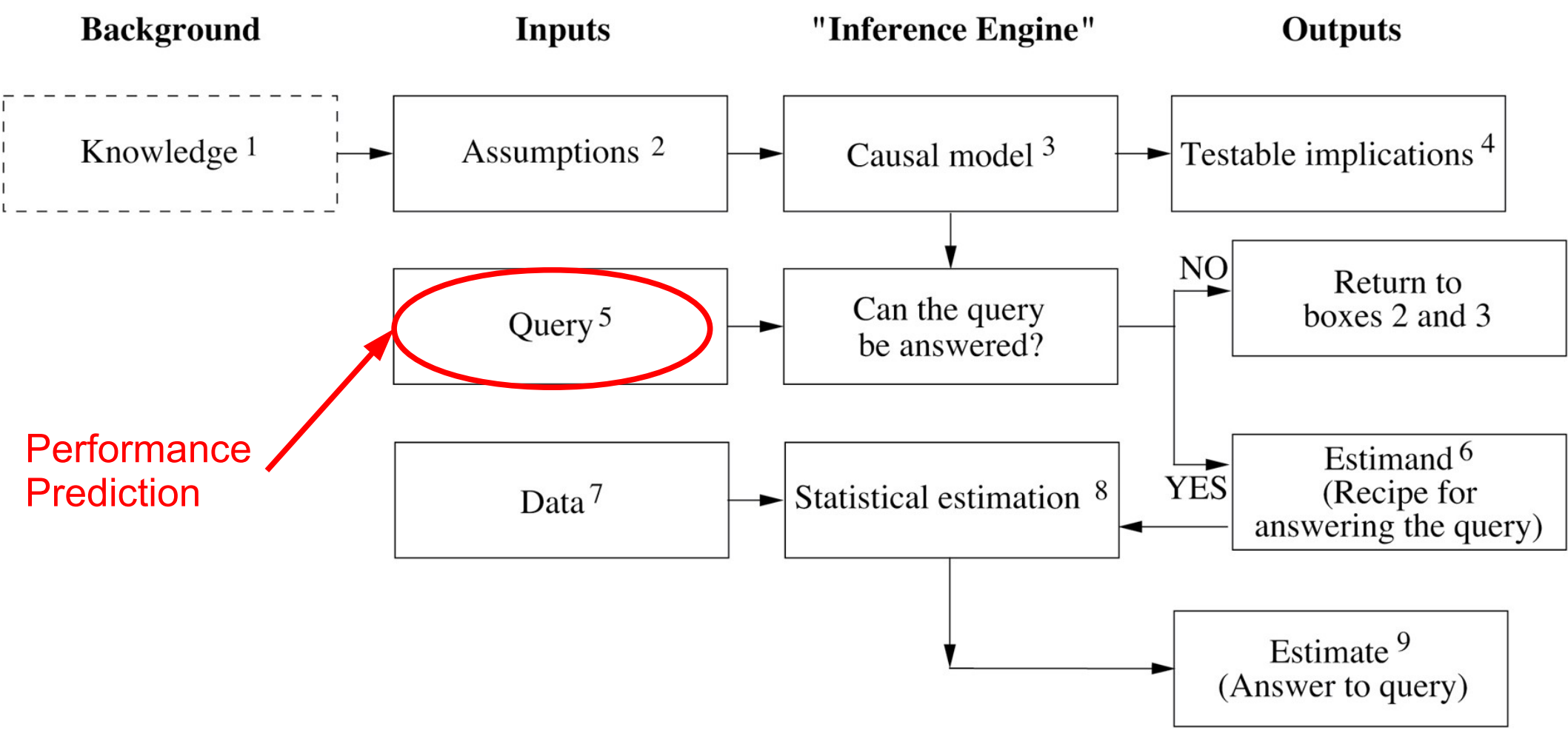
Current machine learning approaches

Causal Inference Networks

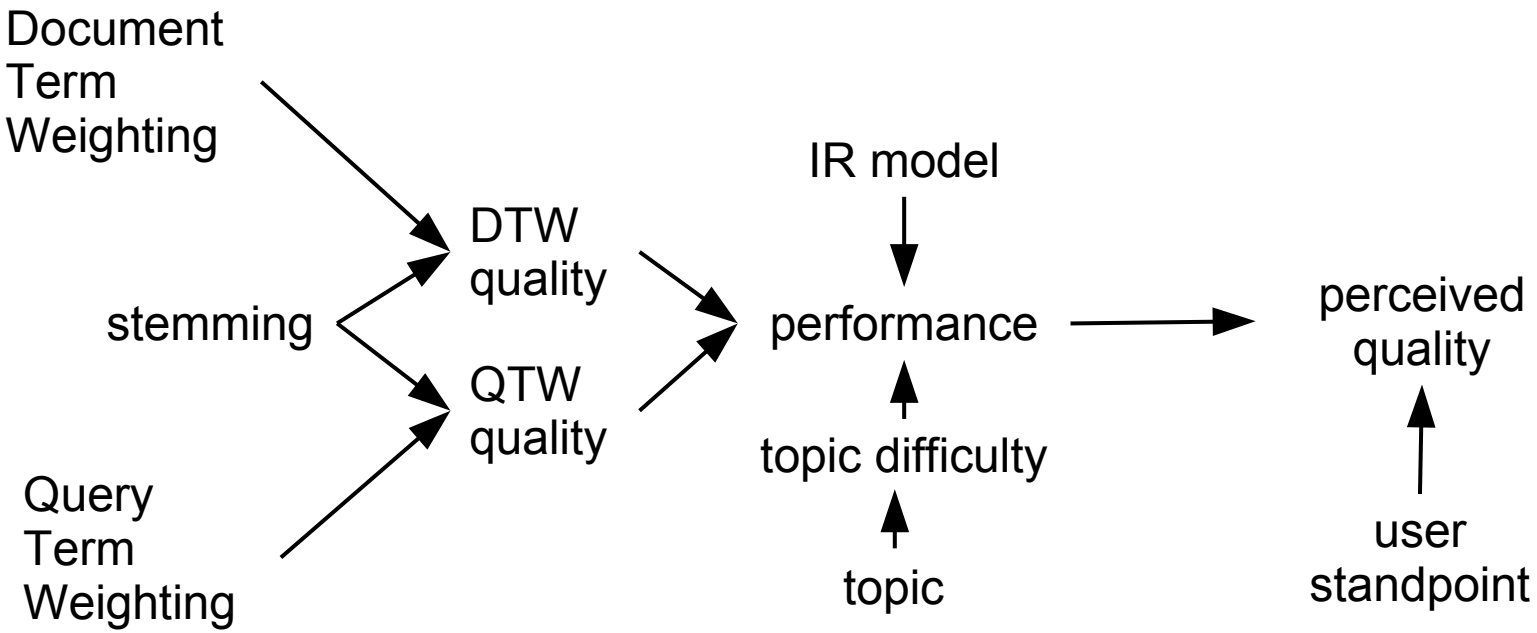




Pearl's Causal Inference Engine



Towards Causal Networks for IR



Conclusion

- We are lacking a multi-stage model of performance
- We have no clue how methods interact
- Effect of collections (Documents x Queries) unexplored
- Causal inference modeling might be a way to address these issues