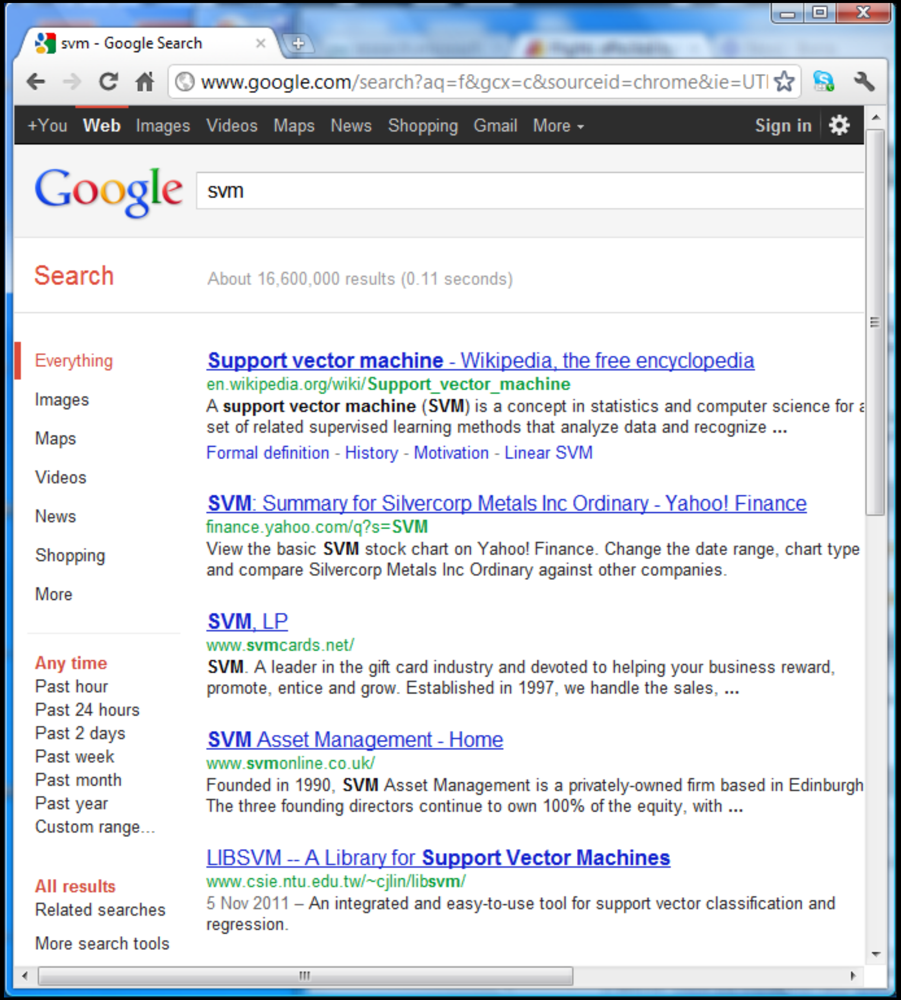
Unbiased Learning-to-Rank with Biased Feedback

Thorsten Joachims, Adith Swaminathan, Tobias Schnabel
International Conference on Web Search and Data Mining, 2017

Presenter: Thorsten Joachims
Dept of Computer Science, Cornell University

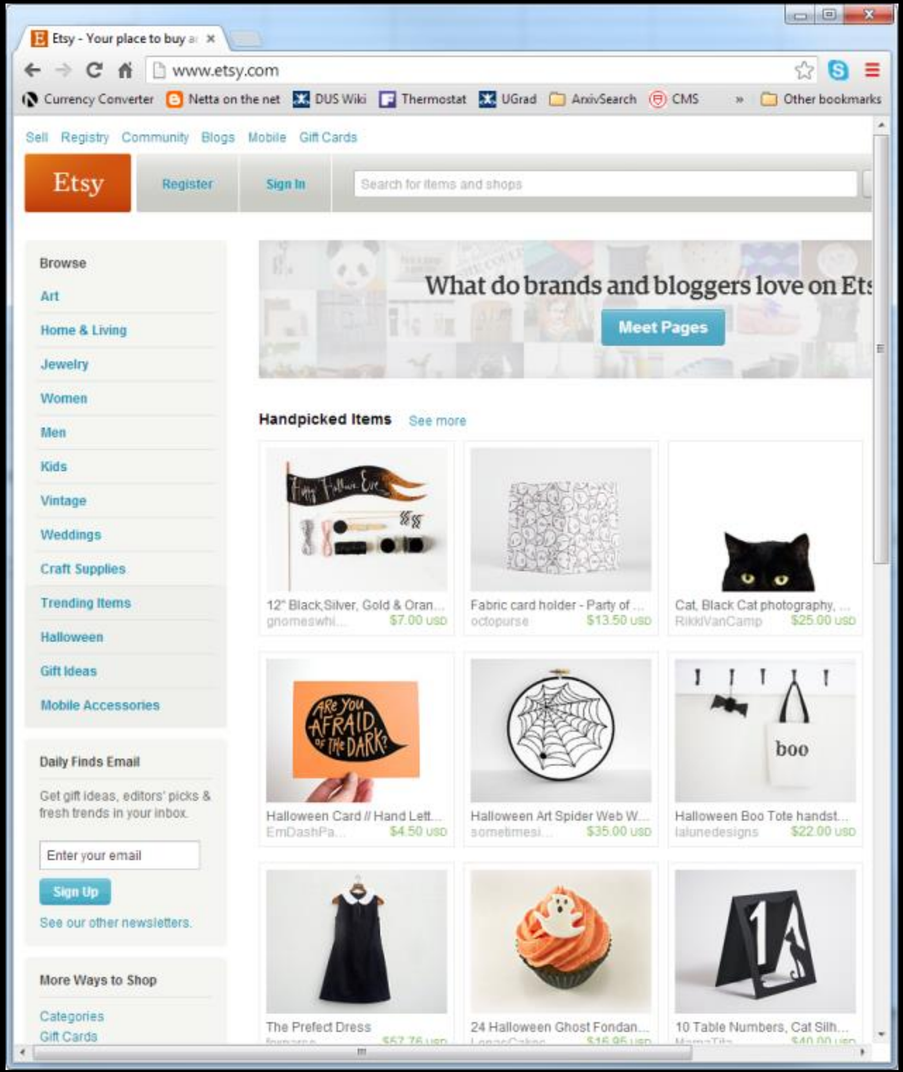
Interaction Logs: Search Engine

- Context x:
  - Query
- Action y:
  - Ranking
- Reward/Loss $\Delta(y|x)$:
  - Search cost
  - Information gained
- Feedback:
  - Clicks on SERP



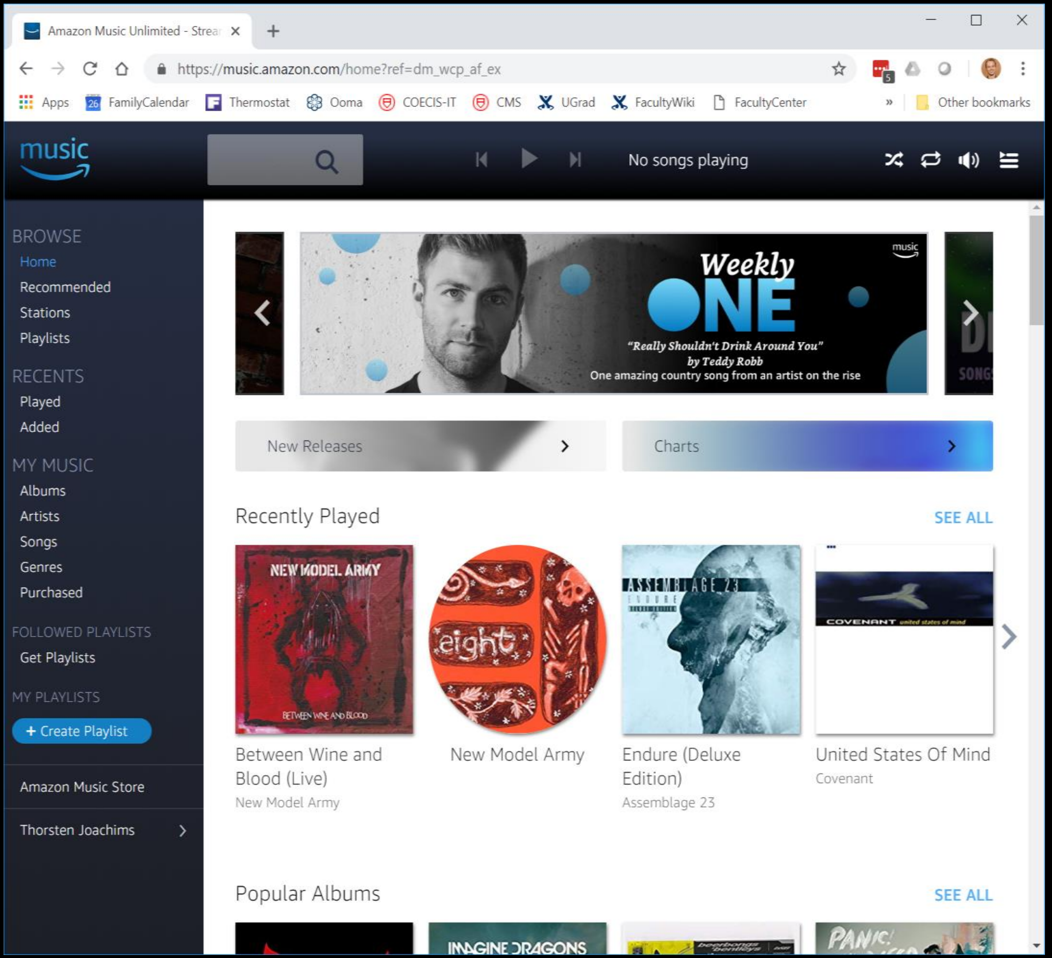
Interaction Logs: Online Retail

- Context x:
  - Category
- Action y:
  - Tile Layout
- Reward/Loss $\Delta(y|x)$:
  - Search cost
  - Product utility
- Feedback:
  - Purchases

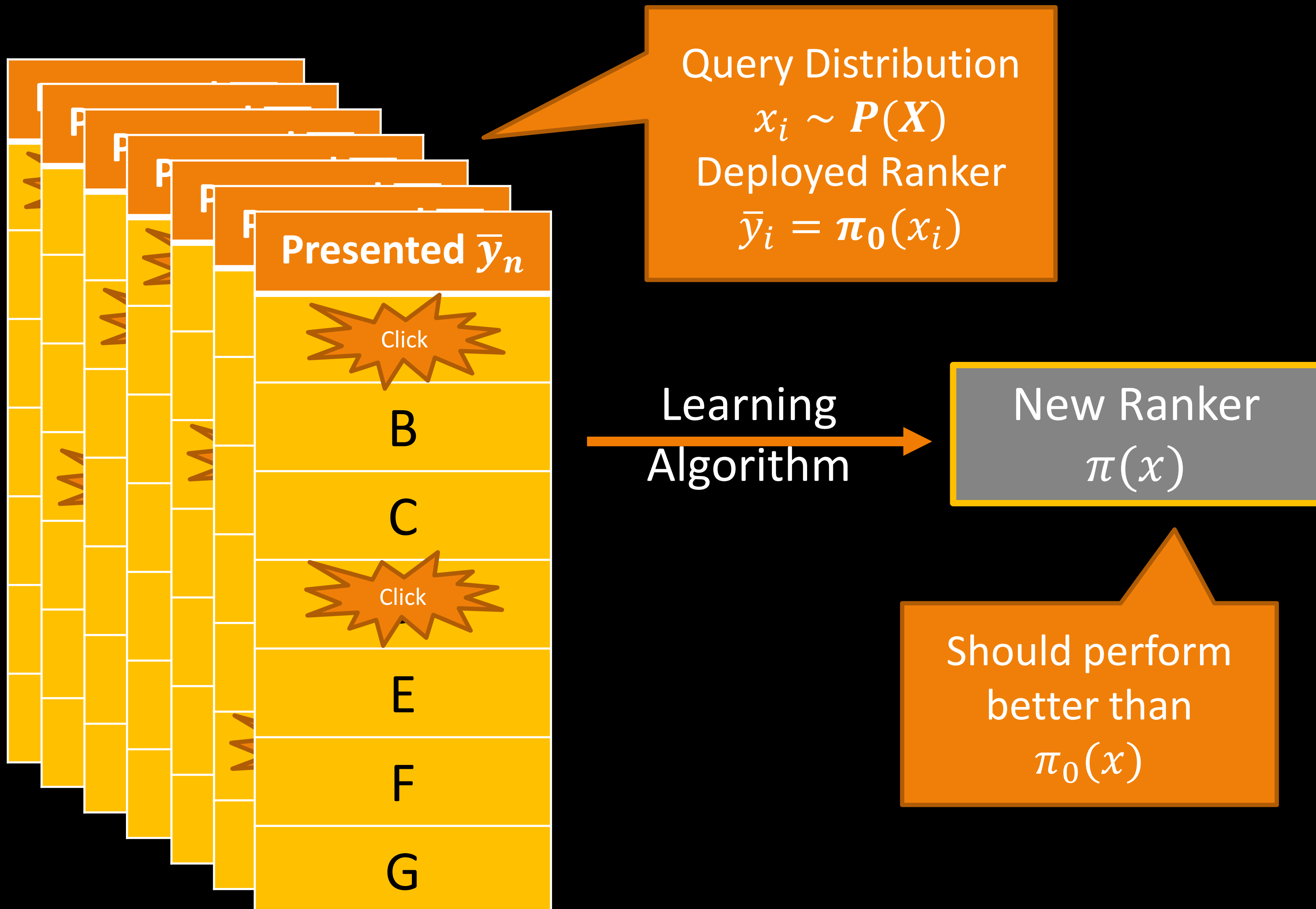


Interaction Logs: Streaming Media

- Context x:
  - User
- Action y:
  - Carousel layout
- Reward/Loss $\Delta(y|x)$:
  - Search cost
  - Enjoyment
- Feedback:
  - Plays

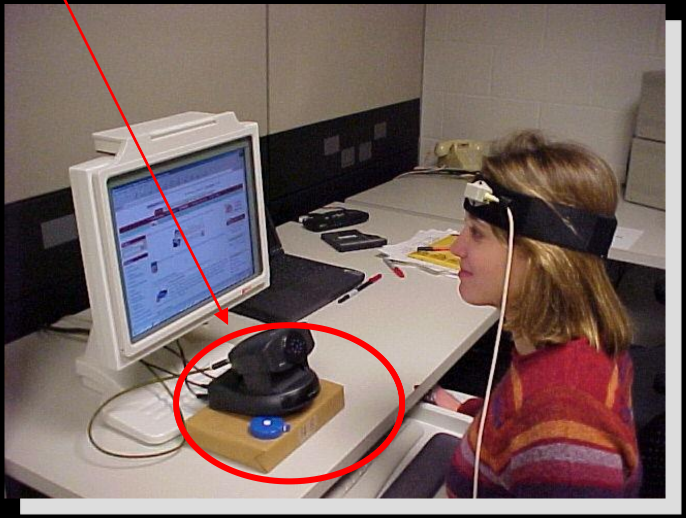


Learning-to-Rank from Clicks



Eye-Tracking

Eye tracking device

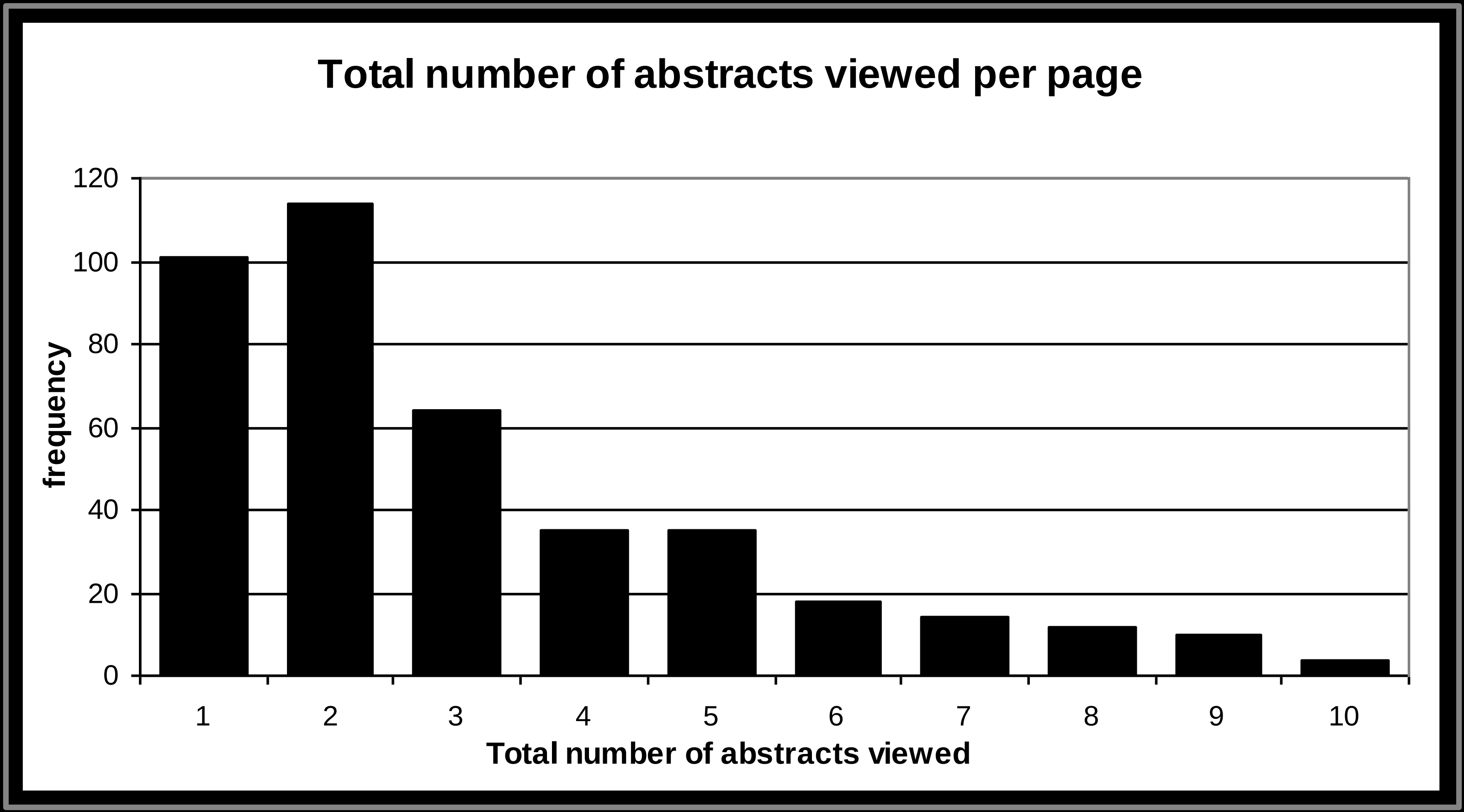


“Scanpath” output depicts pattern of movement throughout screen. Black markers represent fixations.

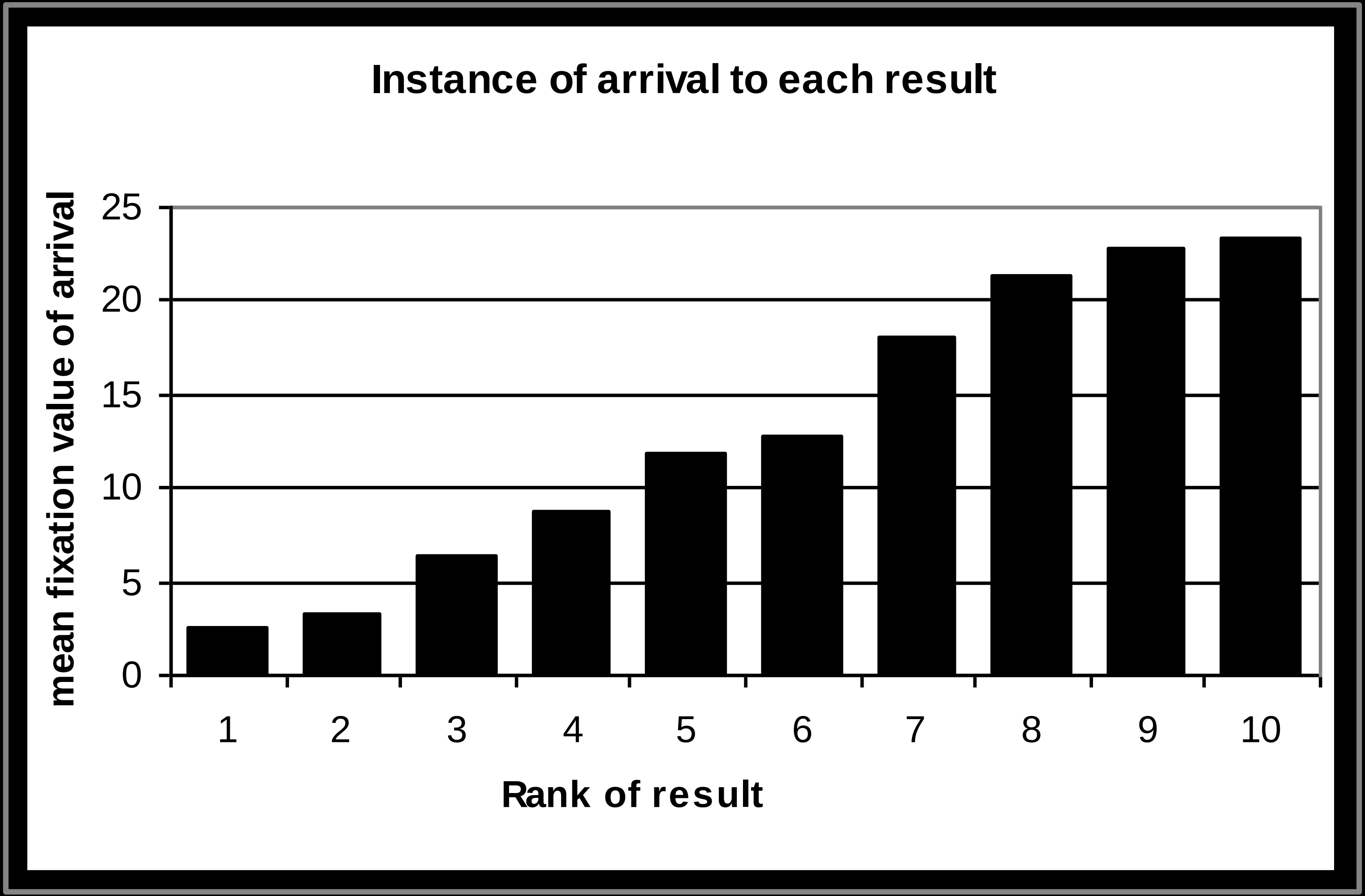
Detect and record where and what people look at

- **Fixations**: ~200-300ms; information is acquired
- **Saccades**: extremely rapid movements between fixations
- **Pupil dilation**: size of pupil indicates interest, arousal

How Many Links do Users View?

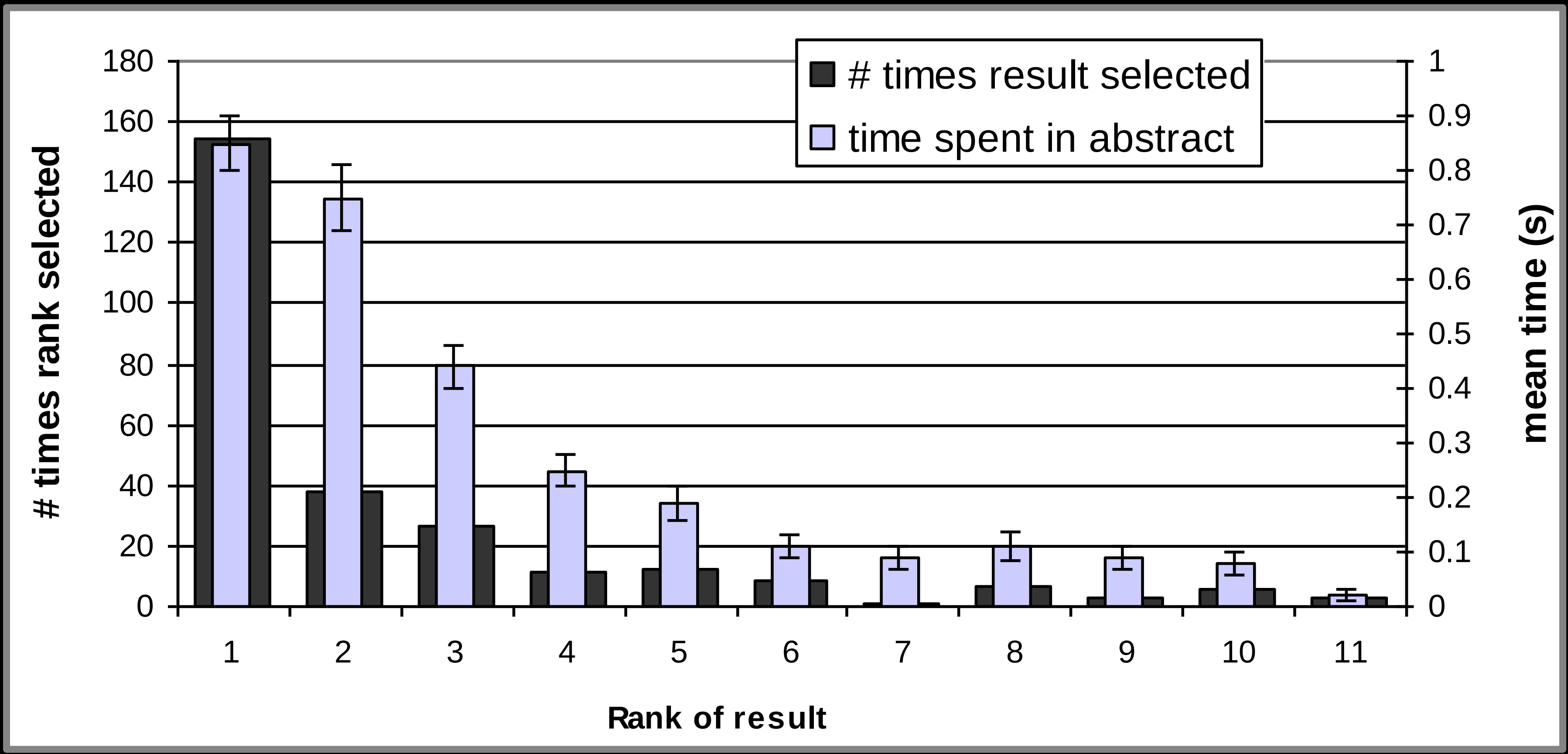


In Which Order are Results Viewed?



=> Users tend to read the results in order

Examination Curve from Eyetracking



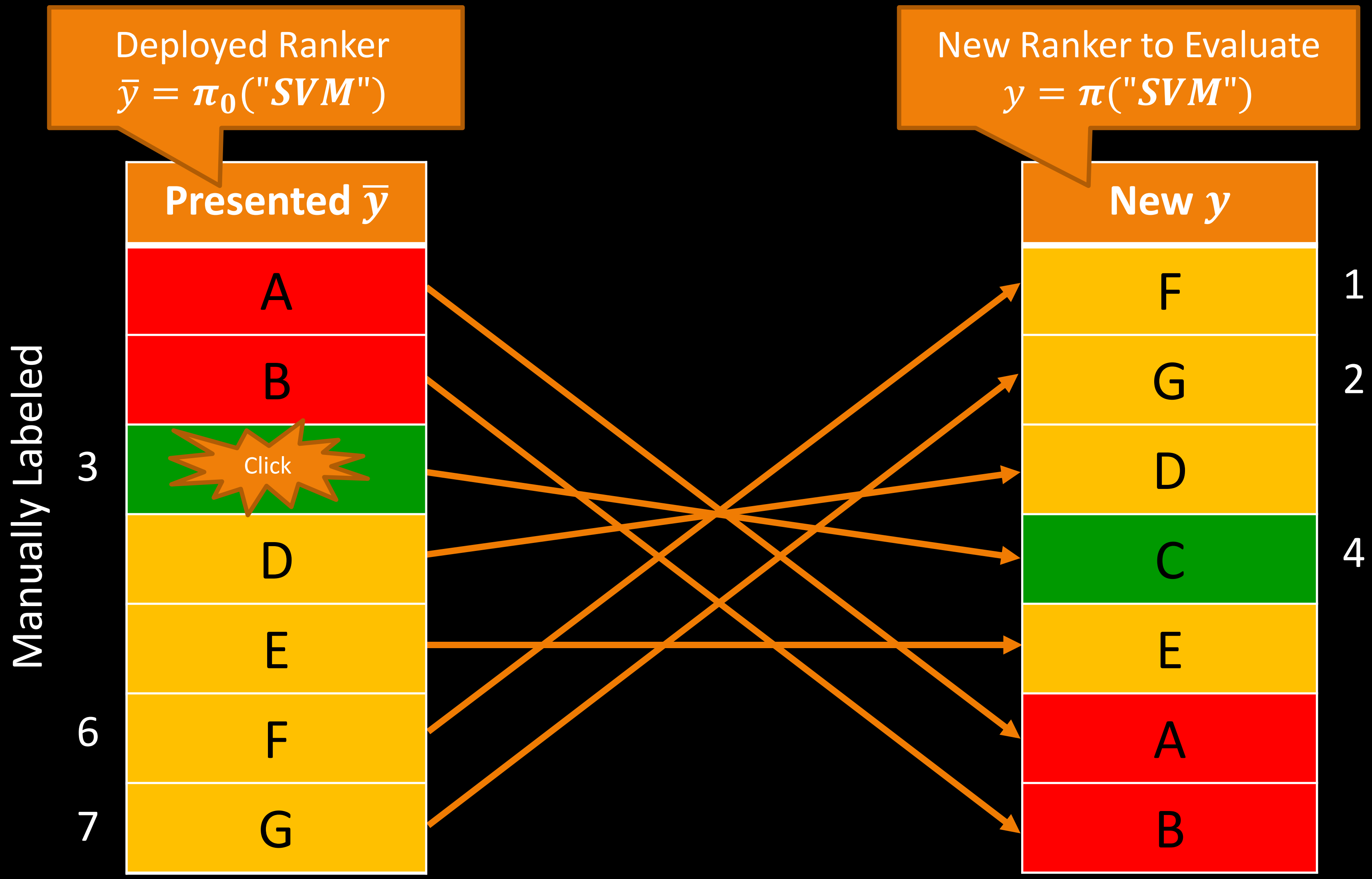
Position → Exposure → Feedback

[Granka et al., 2007]

Outline

- Learning-to-Rank from User Interactions
  – Find new ranking policy π that selects y with better δ
- Batch Learning-to-Rank from Partial Labels
  – Learning from partial and biased feedback
  – Learning Principle: Unbiased Partial-Information ERM
  – Learning Algorithm: Propensity SVM-Rank
- Propensity Estimation for Ranking
  – Break confounding through position randomization
  – Intervention Harvesting
  – Contextual propensity models

Evaluating Rankings



Evaluation with Missing Judgments

- Loss: $\Delta(y|r)$
  - Relevance labels $r_i \in \{0,1\}$
  - This talk: rank of relevant documents

$$\Delta(y|r) = \sum_i rank(i|y) \cdot r_i$$

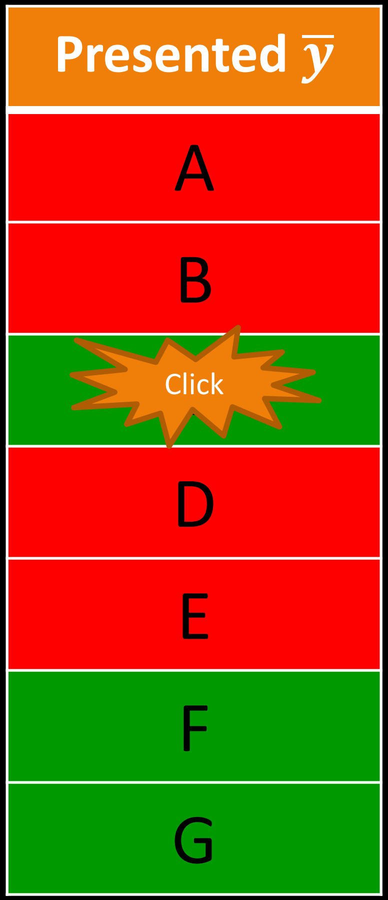
- Assume:
  - Click implies observed and relevant:

$$(c_i = 1) \leftrightarrow (o_i = 1) \wedge (r_i = 1)$$

- Problem:
  - No click can mean not relevant OR not observed

$$(c_i = 0) \leftrightarrow (o_i = 0) \vee (r_i = 0)$$

→ Understand observation mechanism



[Joachims et al., 2017]

Inverse Propensity Score Estimator

- Observation Propensities $Q(o_i = 1|x, \bar{y}, r)$
  - Random variable $o_i \in \{0,1\}$ indicates whether relevance label r_i for is observed

- Inverse Propensity Score (IPS) Estimator:

$$\hat{\Delta}(y|r,o) = \sum_{i:c_i=1} \frac{rank(i|y)}{Q(o_i = 1|\bar{y}, r)}$$

New Ranking

- Unbiasedness: $E_o\left[\hat{\Delta}(y|r,o)\right] = \Delta(y|r)$

Presented $\bar{y}$	Q
A	1.0
B	0.8
C	0.5
D	0.2
E	0.2
F	0.2
G	0.1

[Horvitz & Thompson, 1952] [Rubin, 1983] [Zadrozny et al., 2003] [Langford, Li, 2009] [Joachims et al., 2017]

ERM for Partial-Information LTR

- Unbiased Empirical Risk:

$$\hat{R}_{IPS}(\pi) = \frac{1}{N} \sum_{(x,\bar{y},c)\in S} \sum_{i:c_i=1} \frac{rank(i|\pi(x))}{Q(o_i = 1|\bar{y}, r)}$$

Consistent Estimator of True Performance

- ERM Learning:

$$\hat{\pi} = \underset{S}{\operatorname{argmin}}\left[\hat{R}_{IPS}(\pi)\right]$$

Consistent ERM Learning

- Questions:
  - How do we optimize this empirical risk in a practical learning algorithm?
  - How do we define and estimate the propensity model $Q(o_i = 1|\bar{y}, r)$?

[Joachims et al., 2017]

Propensity-Weighted SVM Rank

- Data: $$S = \left(x_j, d_j, D_j, q_j\right)^n$$

Query

Clicked

Others

Propensity

- Training QP:

$$w^* = \underset{w,\xi\geq 0}{\operatorname{argmin}} \frac{1}{2} w \cdot w + \frac{C}{n} \sum_j \frac{1}{q_j} \sum_i \xi_j^i$$
$$\forall \bar{d}^i \in D_1 : w \cdot \left[\phi(x_1, d_1) - \phi\left(x_1, \bar{d}^i\right)\right] \geq 1 - \xi_1^i$$
$$\vdots$$
$$\forall \bar{d}^i \in D_n : w \cdot \left[\phi(x_n, d_n) - \phi\left(x_n, \bar{d}^i\right)\right] \geq 1 - \xi_n^i$$

Optimizes convex upper bound on unbiased IPS risk estimate!

- Loss Bound:

$$\forall w : rank(d, sort(w \cdot \phi(x, d)) \leq \sum_i \xi^i + 1$$

[Herbrich at al., 1999] [Joachims et al., 2002] [Joachims et al., 2017]

Position-Based Propensity Model

- Model:

$$P\big(c_i = 1 | r_i, rank(i|\bar{y})\big) = q_{rank(i|\bar{y})} \cdot [r_i = 1]$$

- Assumptions
  - Examination only depends on rank
  - Click reveals relevance if rank is examined

Presented $\bar{y}$	Q
A	q_1
B	q_2
C	q_3
D	q_4
E	q_5
F	q_6
G	q_7

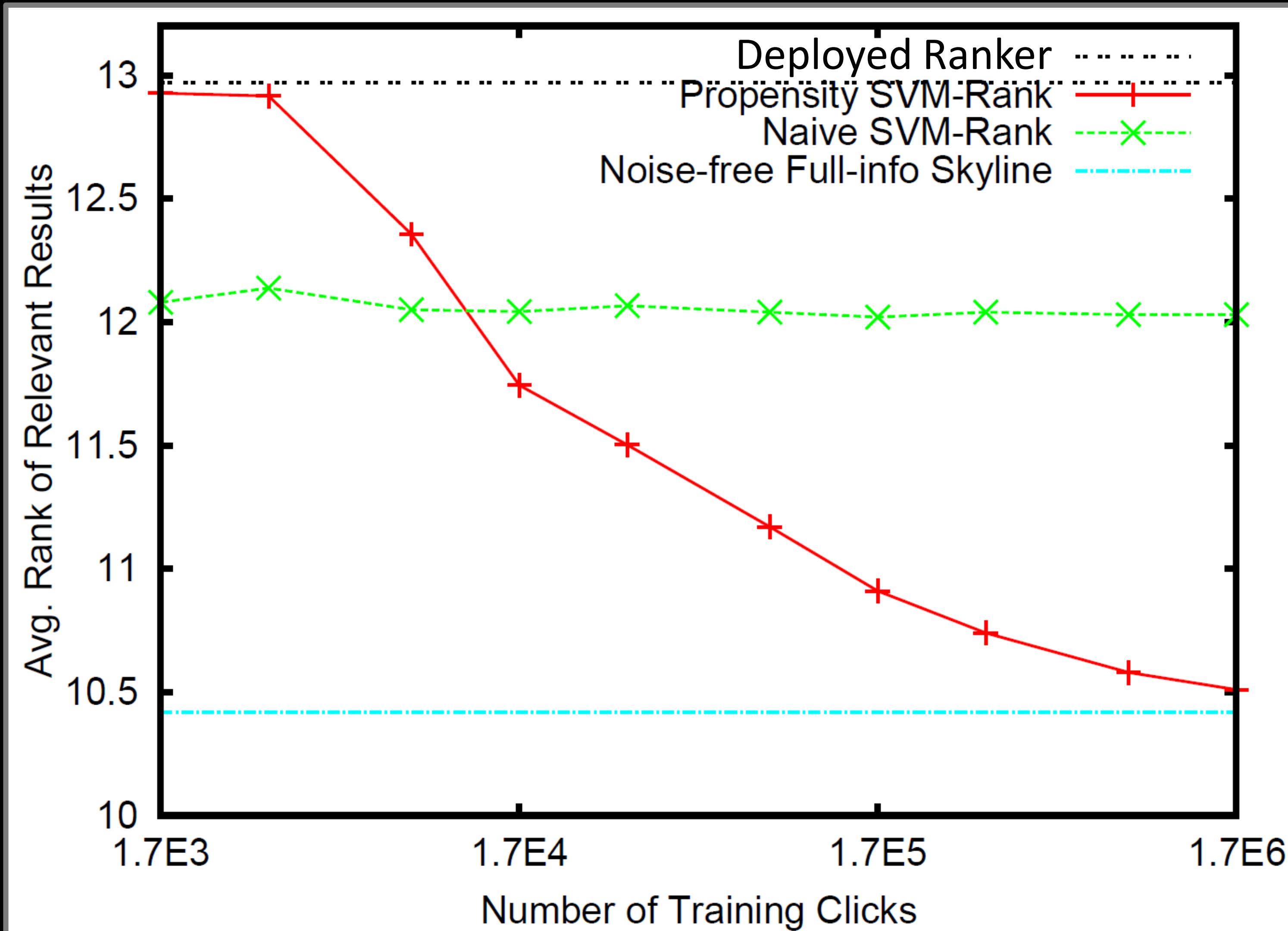
[Richardson et al., 2007] [Chuklin et al., 2015] [Wang et al., 2016]

Experiments

- Yahoo Web Search Dataset
  - Full-information dataset
  - Binarized relevance labels
- Generate synthetic click data based on
  - Position-based propensity model with $q_r = \left(\frac{1}{r}\right)^{\eta}$
  - Baseline "deployed" ranker to generate $\bar{y}$
  - 33% noisy clicks on irrelevant docs

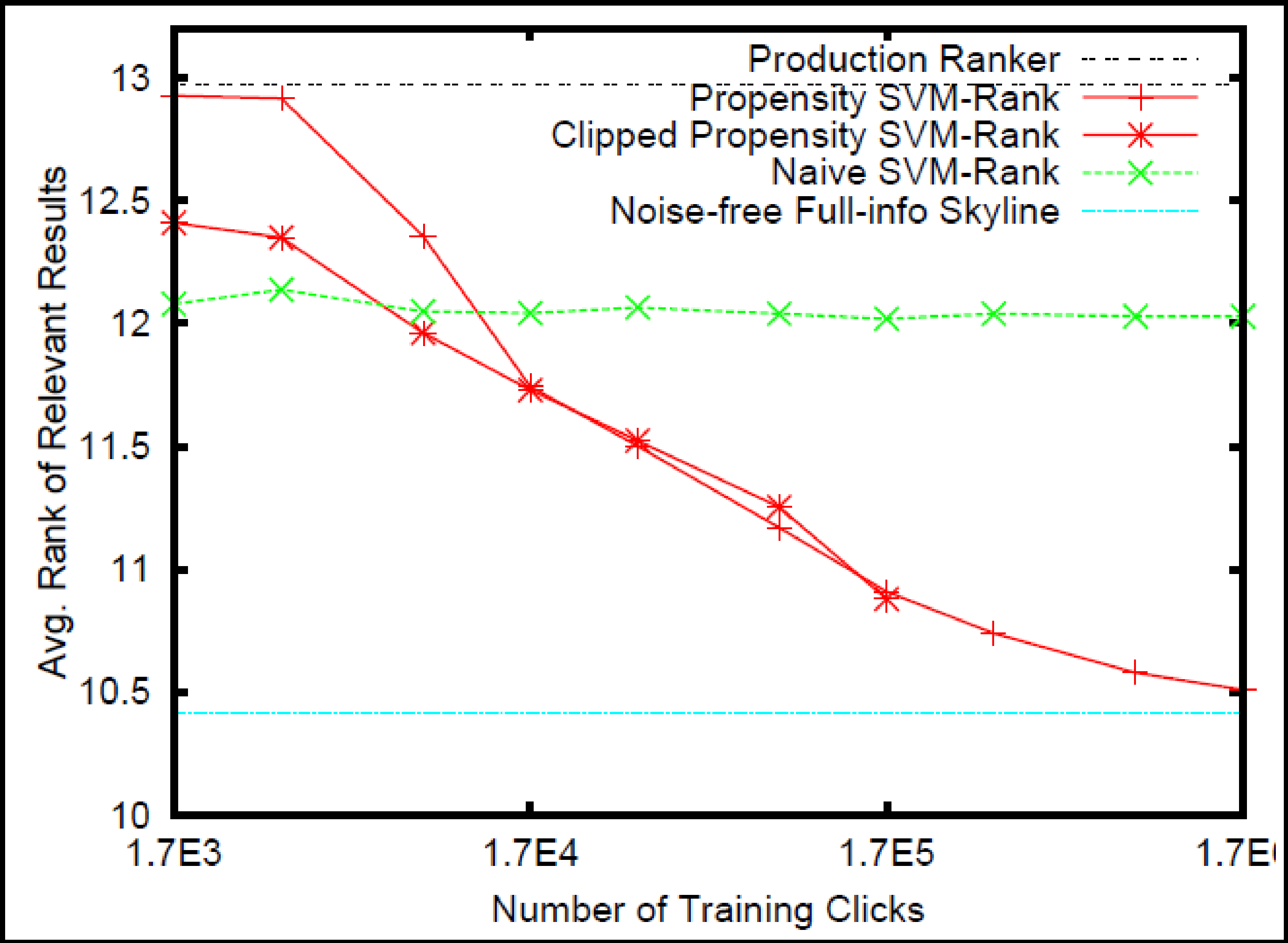
Presented $\bar{y}$	Q
A	q_1
B	q_2
Click	q_3
D	q_4
E	q_5
Click	q_6
G	q_7

Scaling with Training Set Size



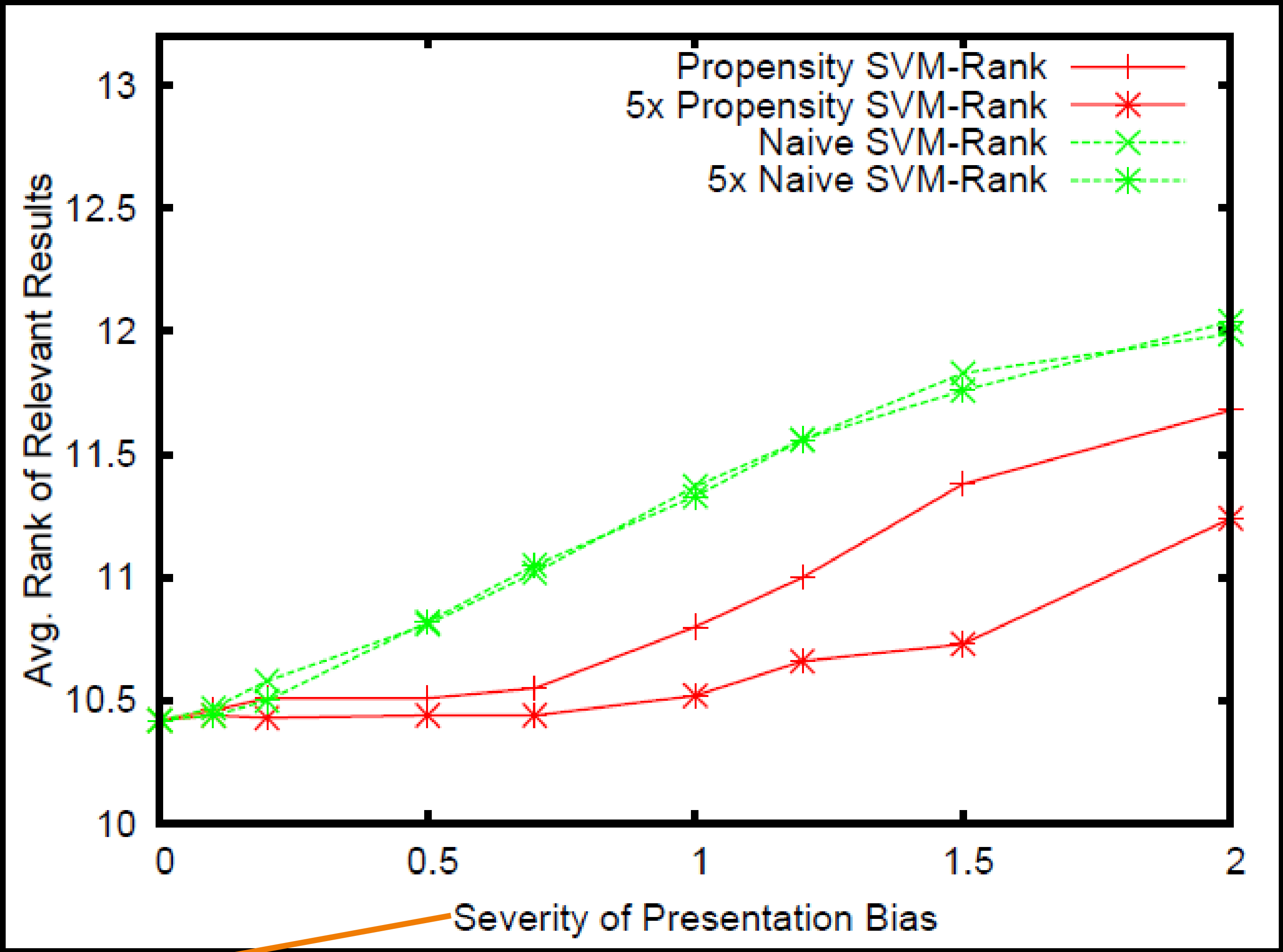
[Joachims et al., 2017]

Scaling with Training Set Size



[Joachims et al., 2017]

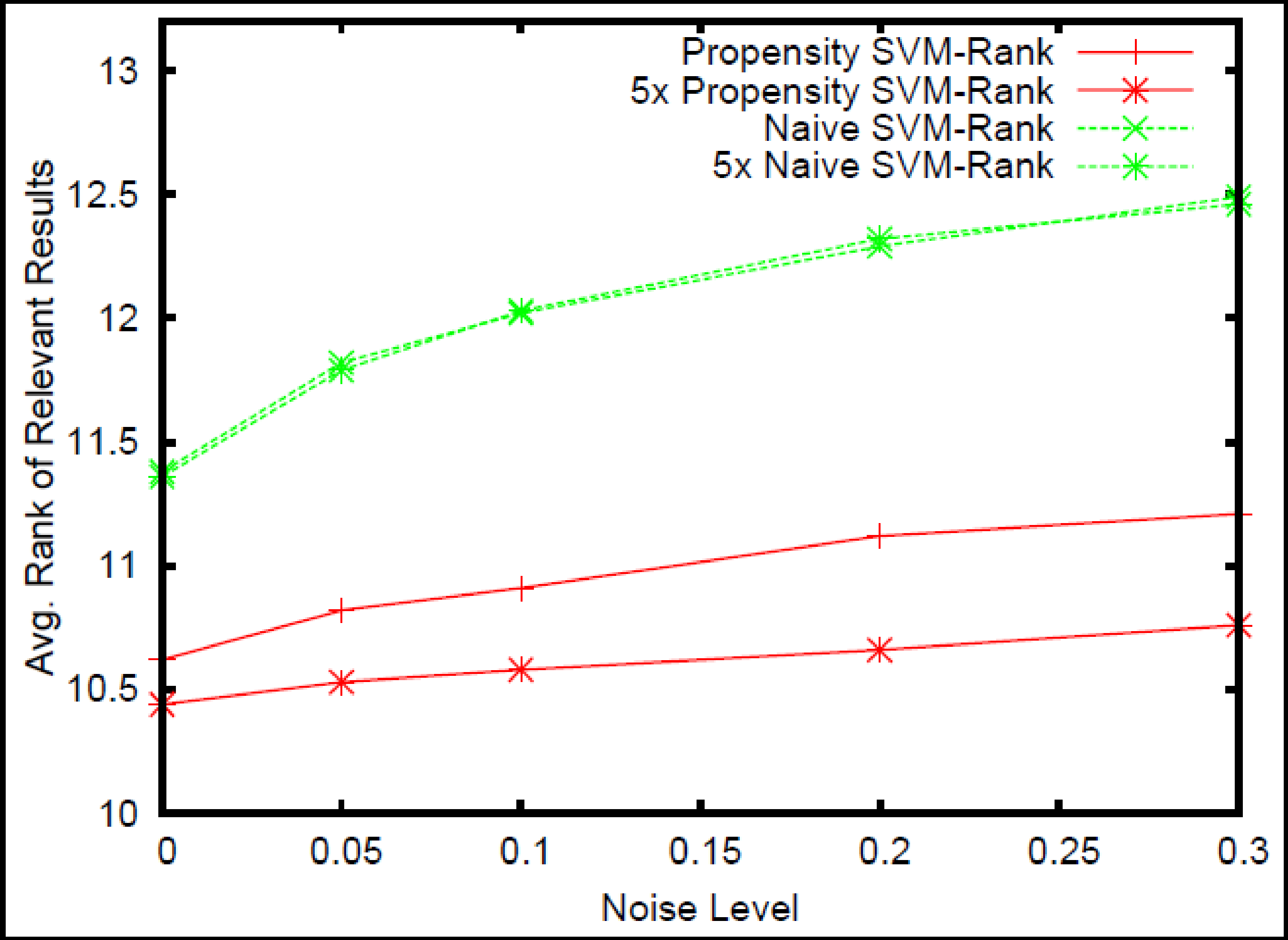
Severity of Presentation Bias



$$q_r = \left(\frac{1}{r}\right)^{\eta}$$

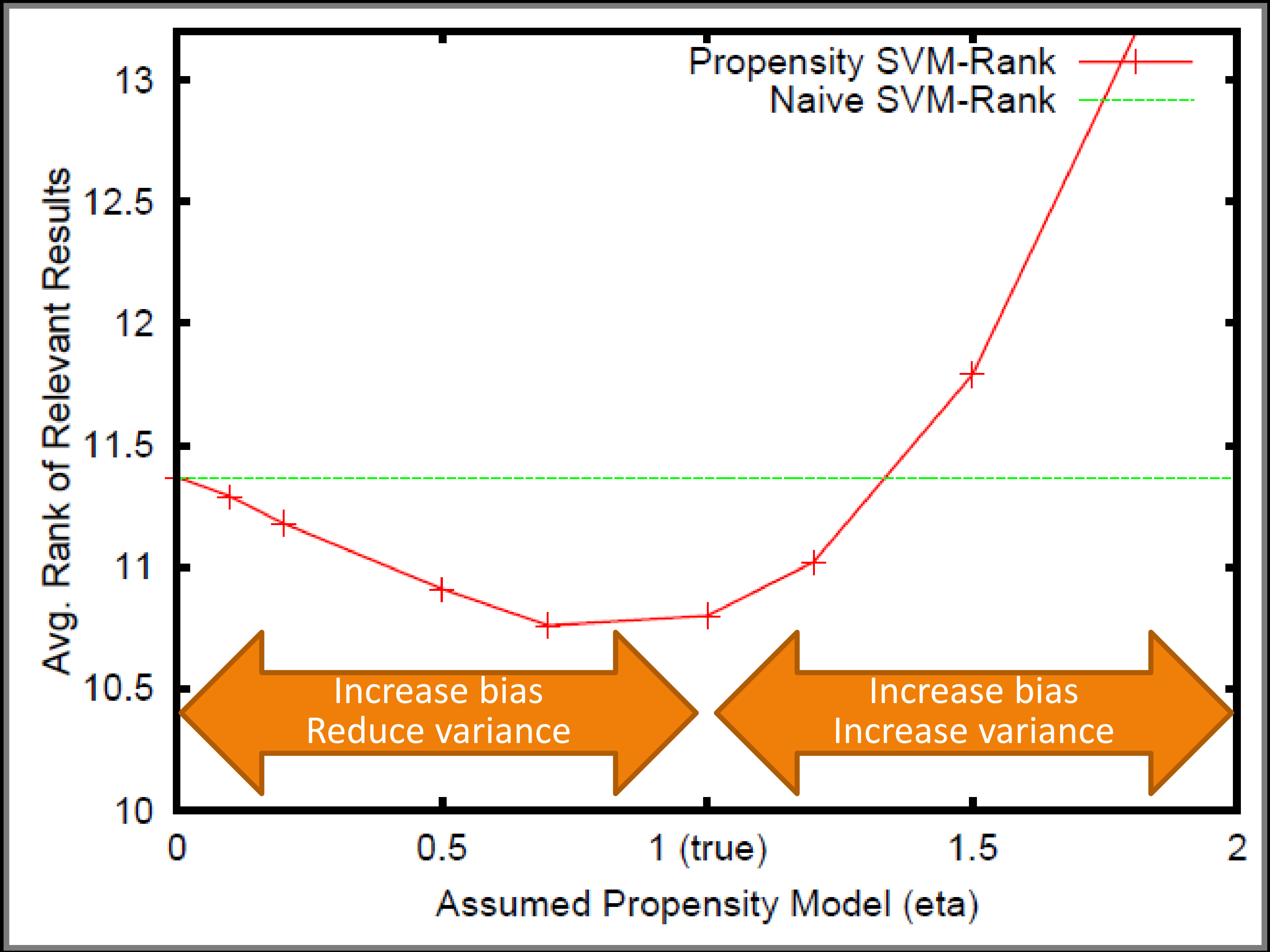
[Joachims et al., 2017]

Increasing Click Noise



[Joachims et al., 2017]

Misspecified Propensities



$$q_r = \left(\frac{1}{r}\right)^{\eta}$$

[Joachims et al., 2017]

Outline

- Learning-to-Rank from User Interactions
  - Find new ranking policy π that selects y with better δ
- Batch Learning-to-Rank from Partial Labels
  - Learning from partial and biased feedback
  - Learning Principle: Unbiased Partial-Information ERM
  - Learning Algorithm: Propensity SVM-Rank
- Propensity Estimation for Ranking
  - Control for relevance through position randomization

Position-Based Propensity Model

- Model:

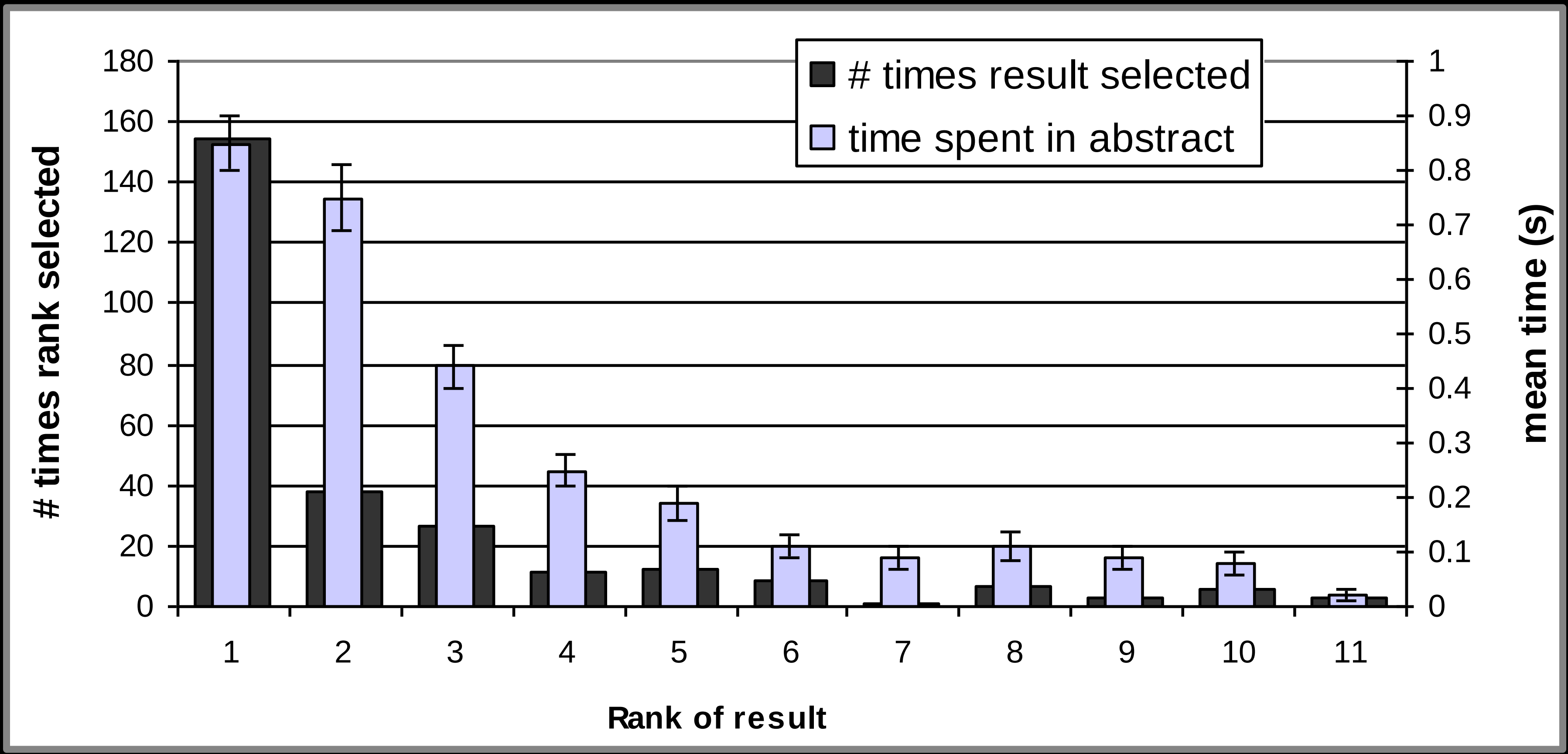
$$P\big(c_i = 1 | r_i, rank(i|\bar{y})\big) = q_{rank(i|\bar{y})} \cdot [r_i = 1]$$

- Assumptions
  - Examination only depends on rank
  - Click reveals relevance if rank is examined

Presented $\bar{y}$	q
A	q_1
B	q_2
C	q_3
D	q_4
E	q_5
F	q_6
G	q_7

[Richardson et al., 2007] [Chuklin et al., 2015] [Wang et al., 2016] [Joachims et al., 2017]

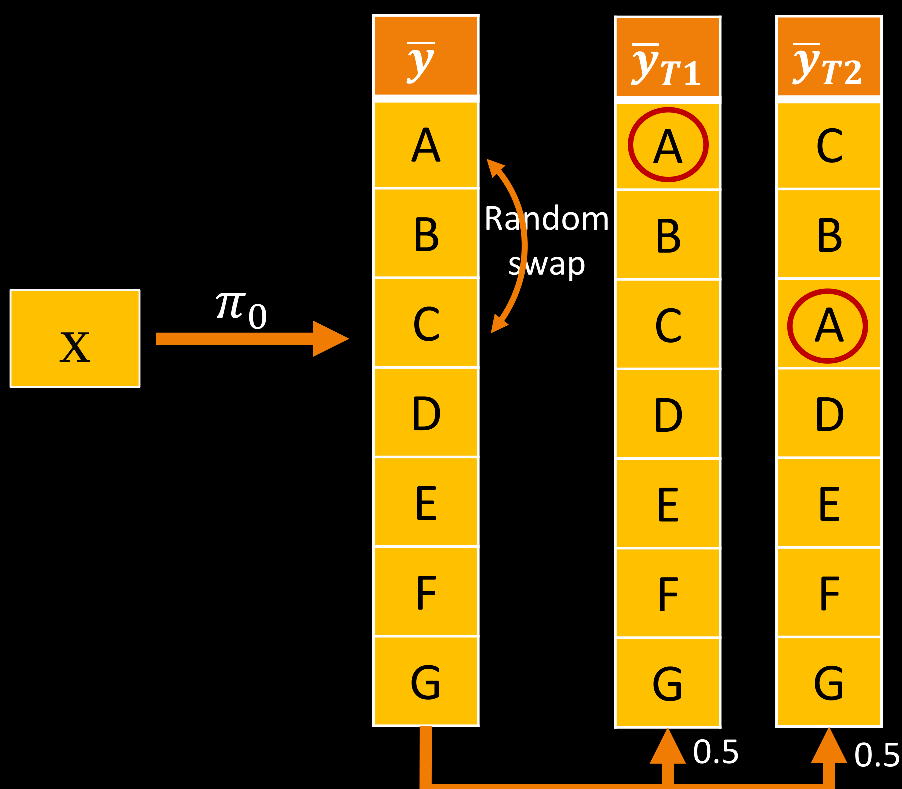
Examination Curve from Eyetracking



[Granka et al., 2007]

Estimating the Propensities

- Idea: Randomization to control for relevance
  → Swap Interventions



$$\mathrm{E}(c_1|T1) = q_1 \cdot E(r_1 = 1|rank(d|\bar{y}) = 1)$$

$$\mathrm{E}(c_3|T2) = q_3 \cdot E(r_1 = 1|rank(d|\bar{y}) = 1)$$

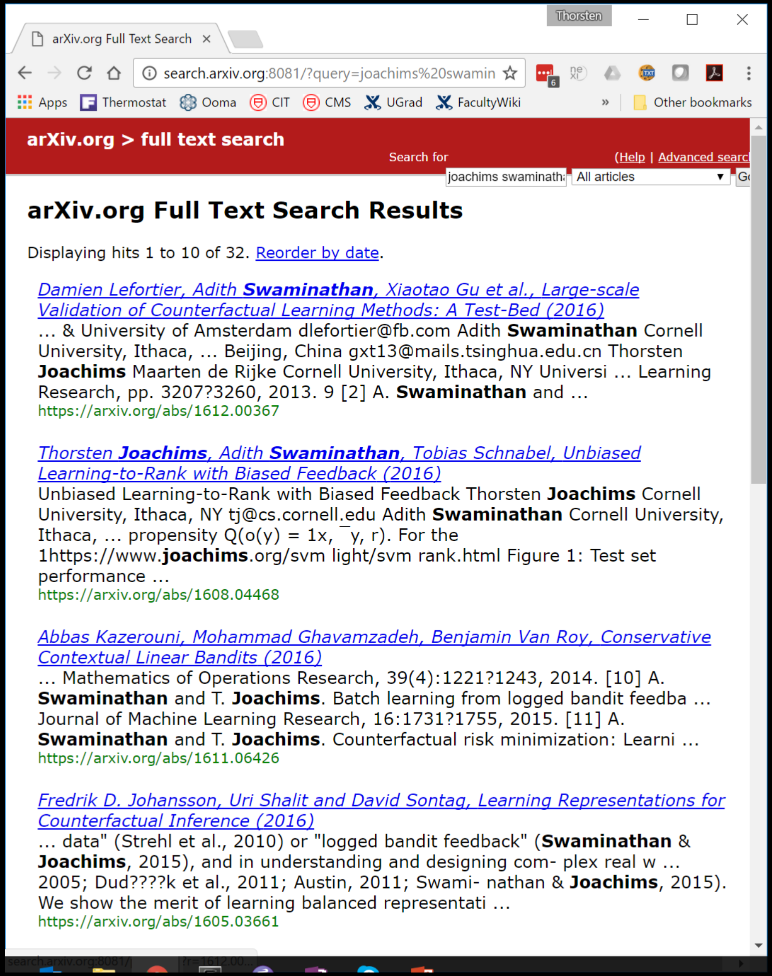
$$\rightarrow \frac{q_1}{q_k} = \frac{\mathrm{E}(c_1|T1)}{\mathrm{E}(c_k|T2)}$$

[Wang et al., 2016; Joachims et al., 2017]

Real-World Experiment

- Arxiv Full-Text Search
  - Run Swap(1,r) experiment to estimate q_r
  - Collect training clicks using production ranker
  - Train naïve / propensity SVM-Rank (1000 features)
  - A/B tests via interleaving

Interleaving Experiment	Propensity SVM-Rank wins	Propensity SVM-Rank loses	Propensity SVM-Rank ties
against Prod	87	48	83
against Naive SVM-Rank	95	60	102



Conclusions and Discussion

- Learning to Rank from User Interactions
- Batch Learning-to-Rank from Partial Labels
  - Find new ranker π that selects y with improved rank metric
  - Positive-only feedback on subset of items
  - Correct for biased feedback due to bias in user exposure
  - Estimate propensities by controling for relevance through swap interventions
- What is still missing?
  - Improve on simplistic propensity model
  - How to deal with zero propensities
  - Biases that do not work through exposure (e.g. Trust Bias)
  - Other learning algorithms and ranking metrics
  - Etc.