THE PERILS OF EXPLORATION UNDER COMPETITION: A COMPUTATIONAL MODELING APPROACH

GUY ARIDOR
COLUMBIA ECONOMICS

KEVIN LIU
COLUMBIA CS

ALEX SLIVKINS
MICROSOFT RESEARCH, NYC

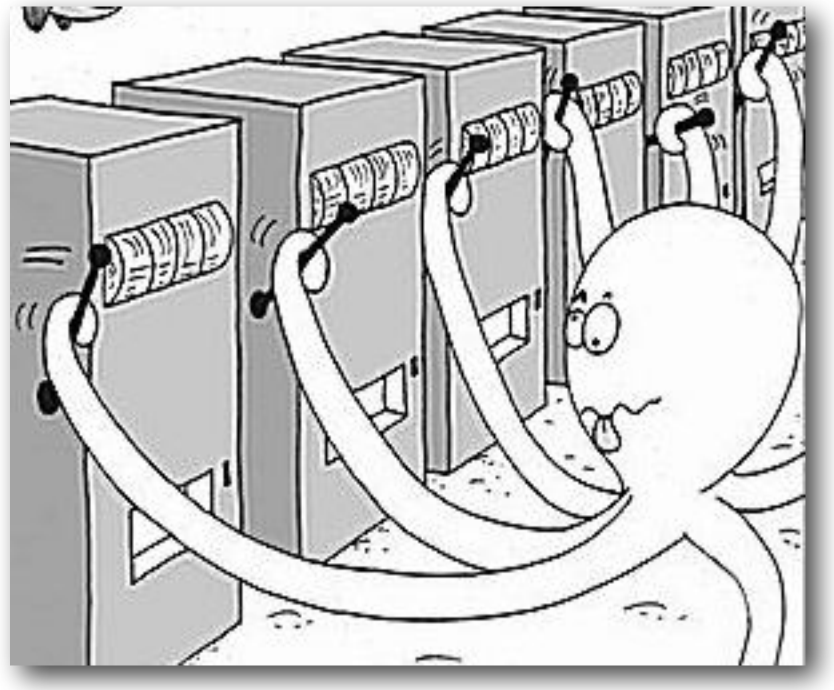
STEVEN WU
UNIVERSITY OF MINNESOTA CS

20TH ACM CONFERENCE ON ECONOMICS AND COMPUTATION

MOTIVATION

- Online platforms increasingly engage in ***product experimentation***
  - Search Engines
  - Recommender Systems
  - E-commerce platforms

Optimizely



- However, they also simultaneously ***compete for users***

- This paper: Firms **compete** for customers and **learn** from the data generated by them

SUMMARY

- Study the tradeoff between ***exploration*** and ***competition.***

  1. Need to incentivize consumers to choose me over competition **today**

  2. Need to explore to gain information to have a better product **tomorrow**

- <u>Questions we study:</u>

  - Does competition incentivize adoption of better algorithms?

  - What is the role that data and reputation play as barriers to entry?

RELATED LITERATURE

- **Multi-armed bandits**
    - Overview (Bubeck and Cesa-Bianchi, 2012), i.i.d rewards (Auer, Cesa-Bianchi, Fischer 2002)
    - Applications in Economics (Bergemann-Välimäki 2006)
- **Tradeoff between exploration, exploitation, and incentives**
    - Information revelation controlled by agents
    - Kremer-Mansour-Perry (2014), Mansour-Slivkins-Syrgkanis (2016), Che-Hörner (2018)
- **Competition vs Innovation**
    - Inverted-U relationship
    - Schumpter (1942), Aghion et.al (2005)
- **Exploration vs Competition (Mansour, Slivkins, Wu 2018)**
    - In their model, consumers do not see signal about firms' past performance
    - Vary competition by consumer response model - here we vary timing of entry
    - Focus on "asymptotic" theoretical results - here we look at more realistic, finite timescales

STOCHASTIC MULTI-ARMED BANDITS

- **Standard Problem -** want to maximize cumulative reward over **T** periods.
  - In each period, select an action from a fixed set of actions.
  - Each action generates a reward (only observe the reward from the action selected)
- In each period, can either:
  - **Exploit -** Make the *best decision today* given the current information
  - **Explore** - Make a sub-optimal decision today (w.r.t. current information) in order to gather information and make ***better decisions tomorrow***

MODEL

FIRM PROBLEM

- **Two** firms, both face the same MAB problem.
- Each round, choose one action from a set of **K** actions (arms)
- Reward of each arm *a* is drawn from {0, 1} independently with expectation *μ(a)*
- Means are initially unknown and firms start with no initial information
- Start with T_0 free observations (selected by the MAB algorithm)
- Aim to maximize their **expected market share**

MODEL

CONSUMERS

- Consumers are homogenous, myopic, and nonstrategic.
    - Don’t try to manipulate the firm’s choice of algorithm
    - Live for one period, aim to maximize their current period utility
- Stylized notion of competition
    - Consumers choose firm according to **one dimension** - quality
    - Choose firm that has a higher reputation score
- The **reputation score** is a sliding window average of the rewards of the last **M** consumers that chose this firm

MODEL

OVERVIEW OF TIMING

- The timing of events is as follows:

1. At t=0, the firms commit to a MAB algorithm
2. Firms receive T_0 free observations
3. Starting at t=1, a new consumer arrives (and lives for only one round), and chooses the firm with the higher quality score
4. The firm that is chosen chooses an action according to its MAB algorithm
5. The consumer experiences a reward and reports it to the firms and the future consumers
6. Repeat 3-5 for T rounds

MODEL

MAB ALGORITHMS

- Model firms as committing to a MAB algorithm
- Utilize the well-known distinction between three classes of **MAB learning algorithms.**
- Select a single representative algorithm from each class:
    - **Greedy** - strive to take actions with maximal mean reward, based on the current information. From this class, utilize Dynamic Greedy (DG)
    - **Exploration-Separating -** exploration choices do not use the rewards observed so far. From this class, utilize Dynamic Epsilon-Greedy (DEG)
    - **Adaptive Exploration** - Sway exploration choices towards more promising alternatives. From this class, utilize Thompson Sampling (TS)
- In isolation, **Adaptive Exploration > Exploration-Separating > Greedy**

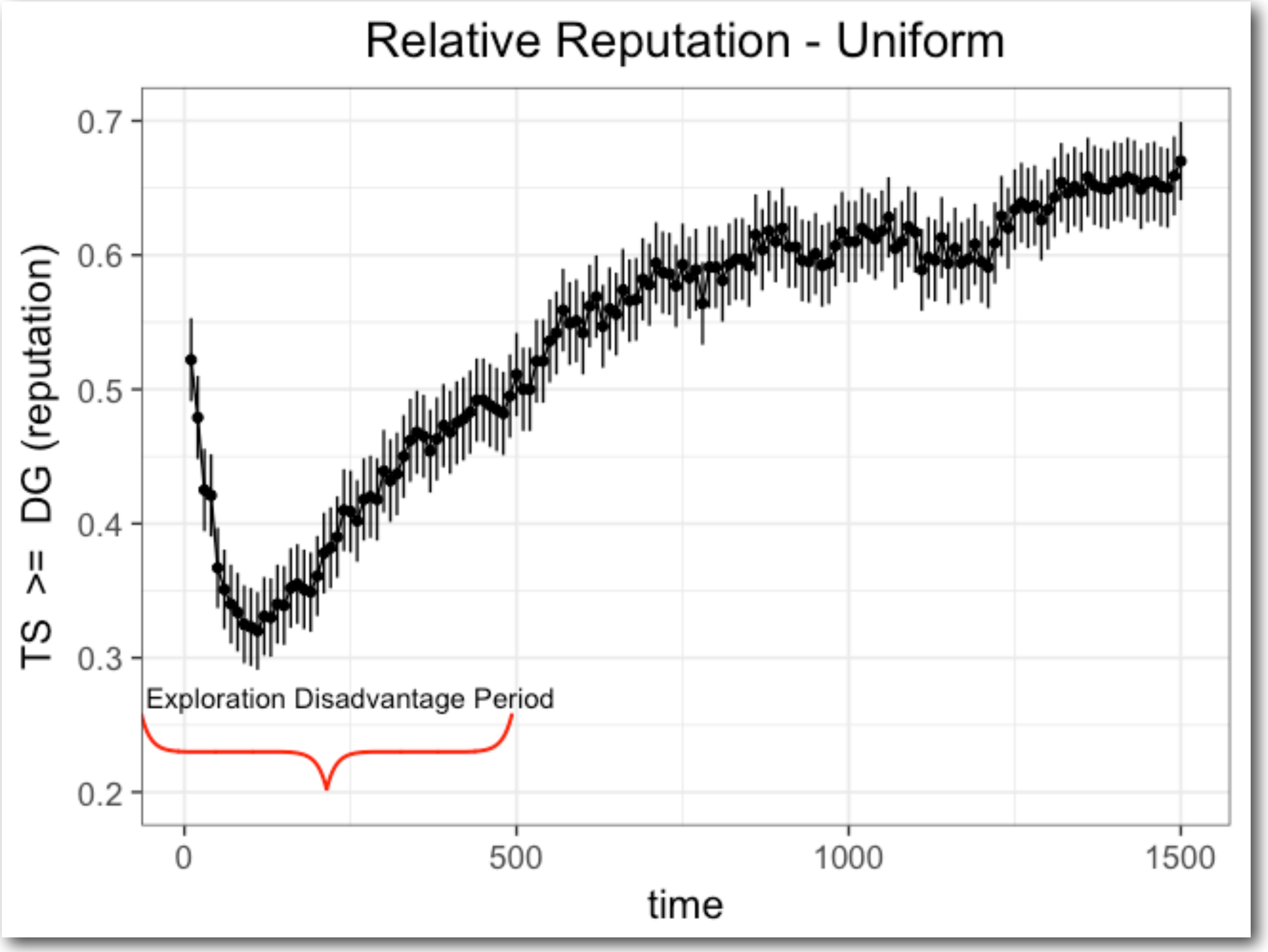
MODEL

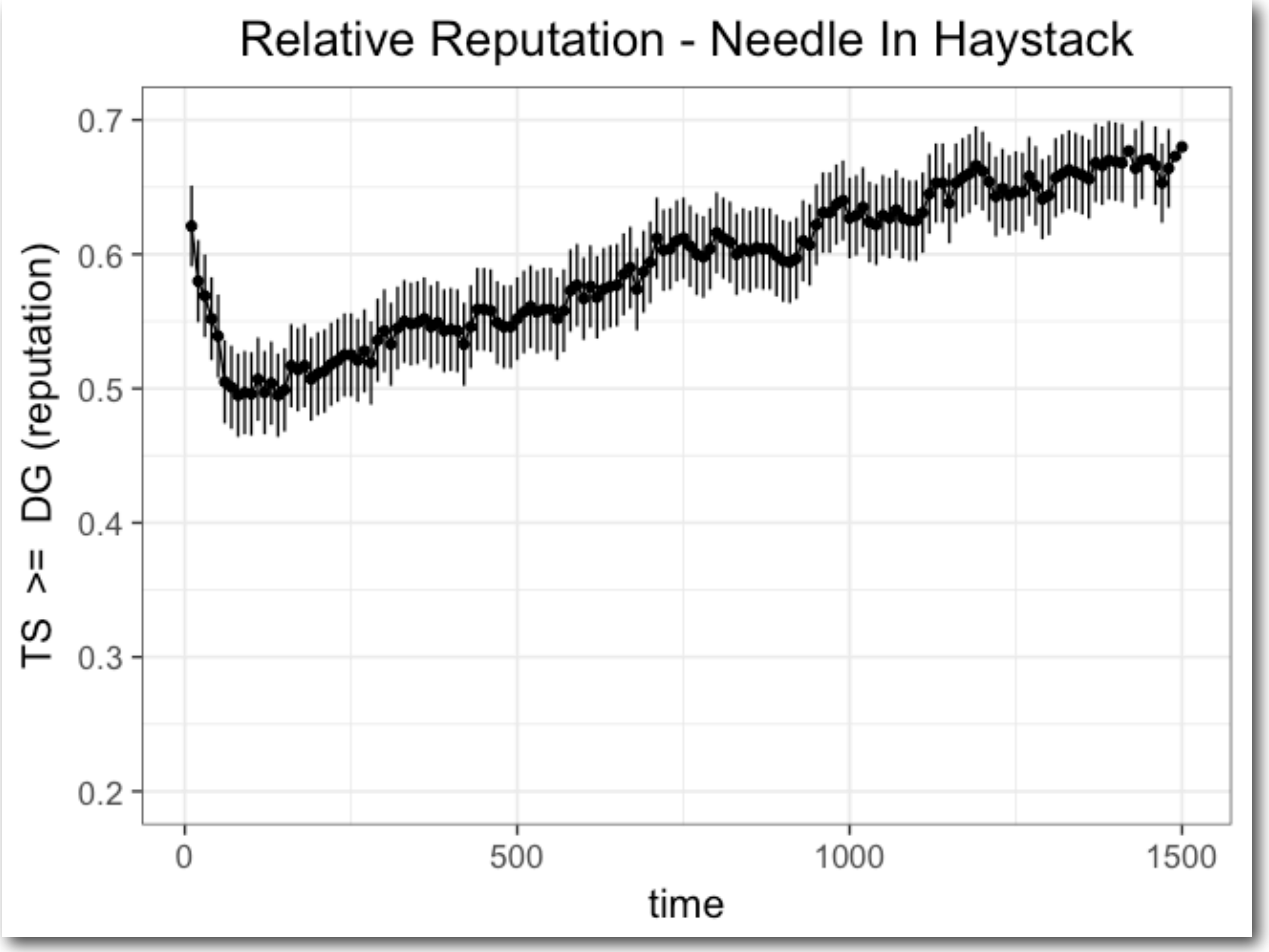
SIMULATIONS

- Study our model via numerical simulation
- We break indifference towards “simpler” algorithms since, in practice, they tend to be easier to deploy (Agrawal et.al 2016, 17)
- Consider three representative classes of instances:
    - *Needle-In-Haystack* - 1 “good” arm, K-1 identical “bad” arms
    - *Uniform* - mean rewards drawn from Uniform[0.25, 0.75]
    - *Heavy Tail* - mean rewards drawn from Beta(0.6, 0.6)

RELATIVE REPUTATION

- Relative reputation - at a given time *t*, what fraction of mean reward vectors in which Alg 1 has a higher quality score than Alg 2





EXPLORATION DISADVANTAGED INSTANCES

- Purposeful exploration can lead to ***short-term reputation consequences***
- Comparing the relative quality score between an algorithm that engages in purpose exploration and the greedy algorithm:
  - When we see an initial decrease in the relative reputation plot below 0.5, we call the instance ***exploration disadvantaged***
- Heavy Tail and Uniform are exploration disadvantaged
- Needle In Haystack is not exploration disadvantaged

COMPETITION INTENSITY LEVELS

- We consider the equilibrium strategies for four separate “competition intensity” levels in our model:

1. “Monopoly”- only one firm in the market

2. “Incumbent” - one firm enters early, is a monopolist for X periods, and then the other firm enters

3. “Permanent duopoly” - both firms enter the market at the same time

4. “Entrant” - the firm enters after there already is an incumbent for X periods

OVERVIEW OF COMPETITION RESULTS

- On exploration disadvantaged instances, we have the following set of results:

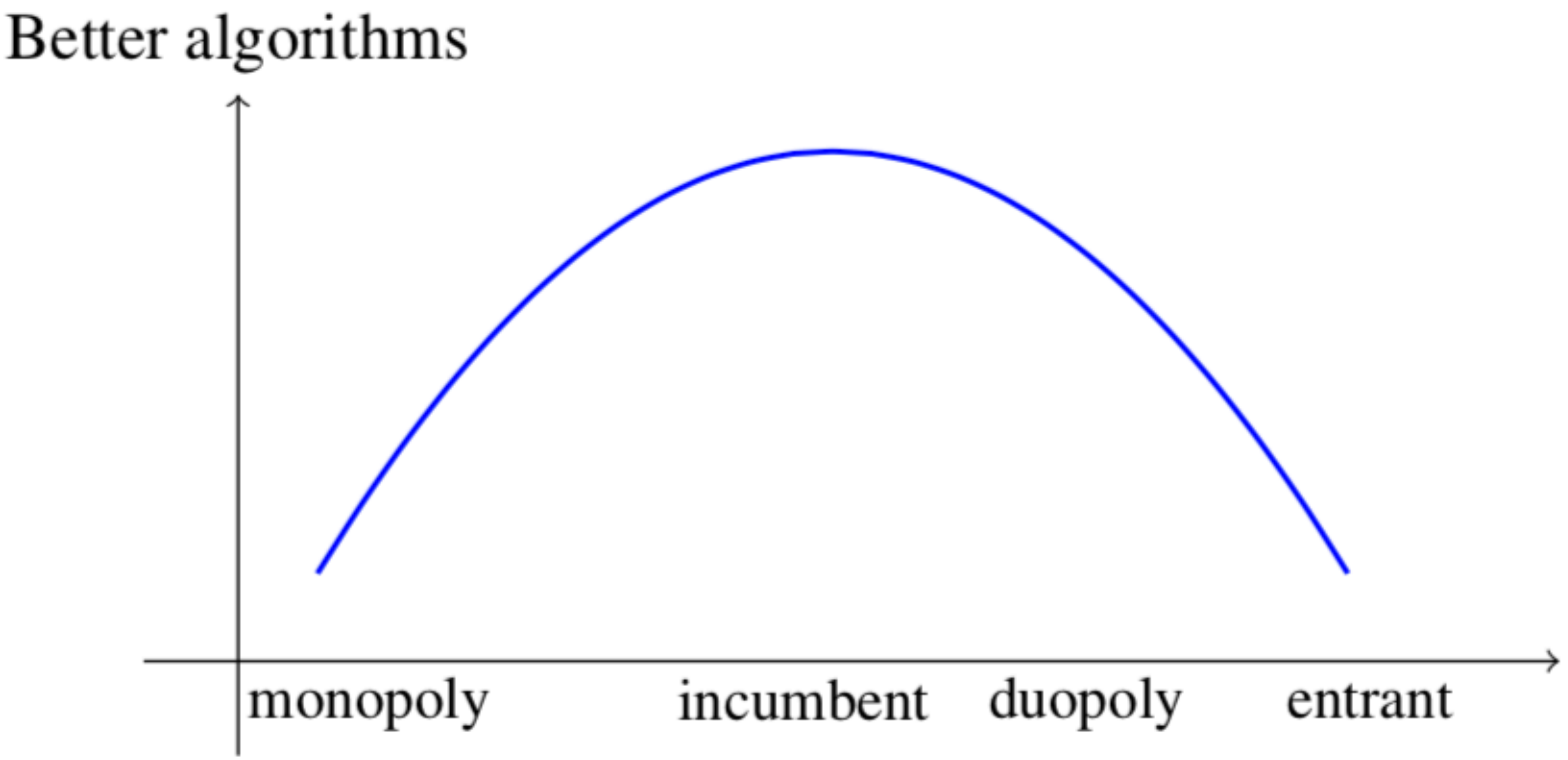


Figure 1: A stylized "inverted-U relationship" between strength of competition and "level of innovation".

PERMANENT DUOPOLY

GAME

- Use simulation to compute expected payoffs

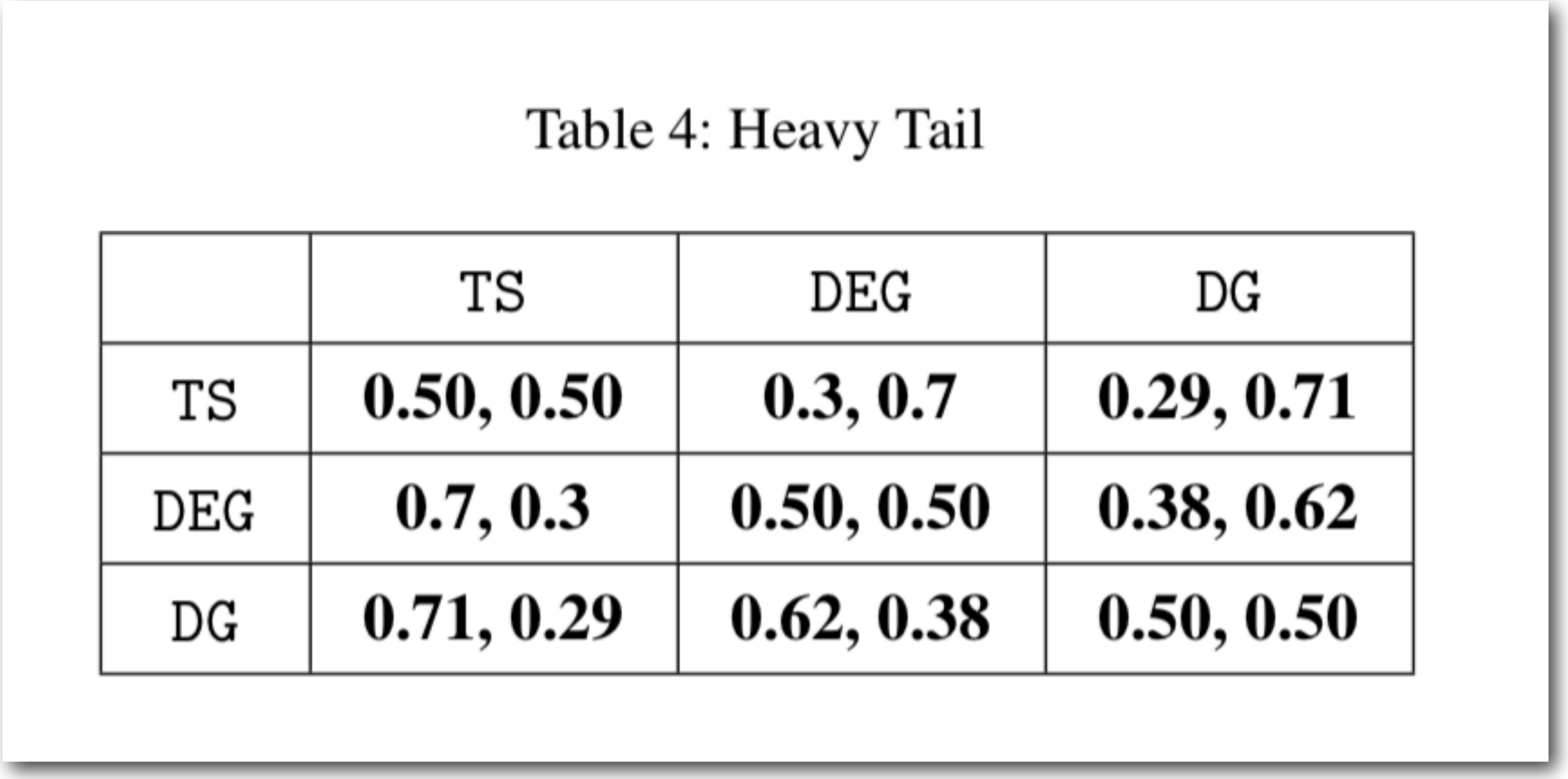


Table 4: Heavy Tail

	TS	DEG	DG
TS	**0.50, 0.50**	**0.3, 0.7**	**0.29, 0.71**
DEG	**0.7, 0.3**	**0.50, 0.50**	**0.38, 0.62**
DG	**0.71, 0.29**	**0.62, 0.38**	**0.50, 0.50**

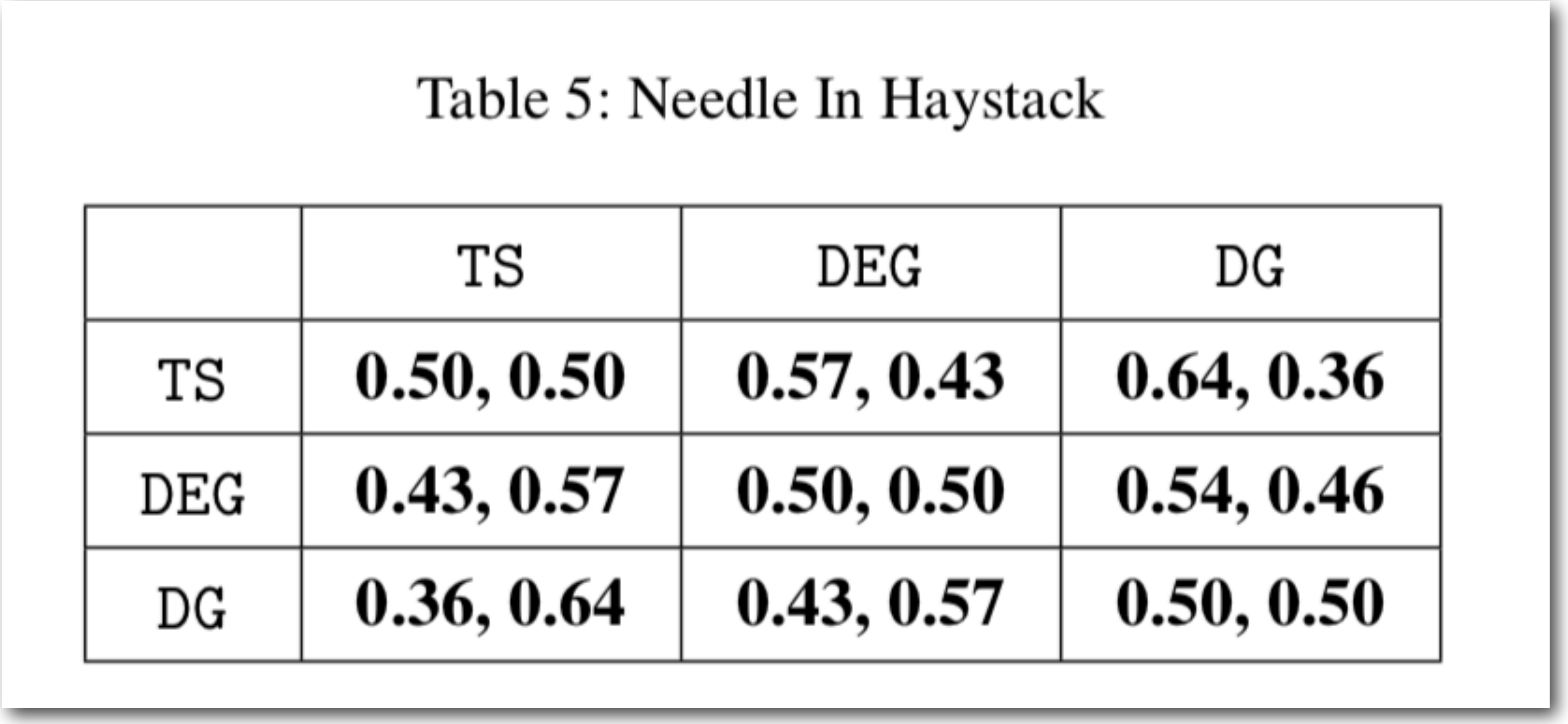
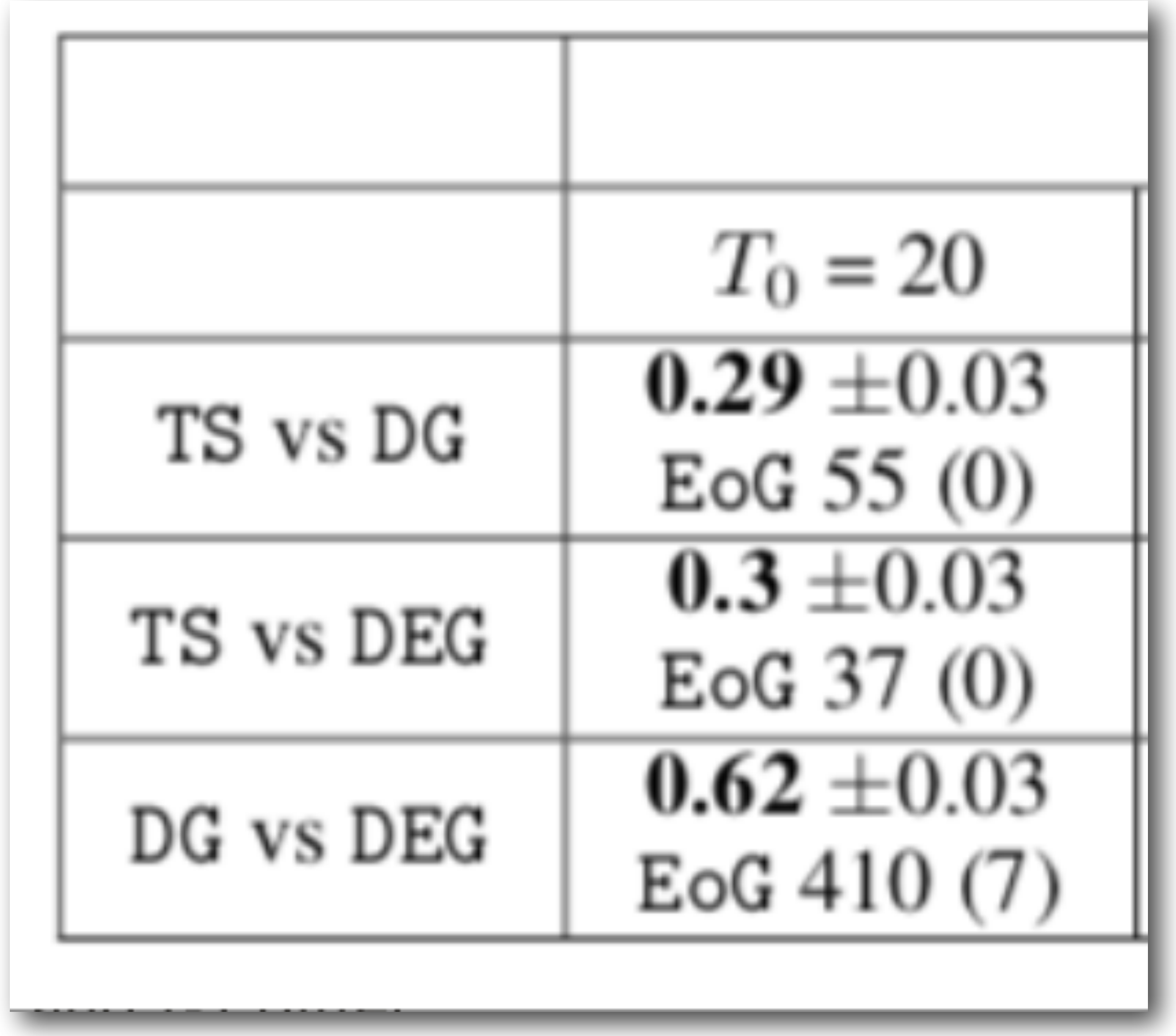


Table 5: Needle In Haystack

	TS	DEG	DG
TS	**0.50, 0.50**	**0.57, 0.43**	**0.64, 0.36**
DEG	**0.43, 0.57**	**0.50, 0.50**	**0.54, 0.46**
DG	**0.36, 0.64**	**0.43, 0.57**	**0.50, 0.50**

PERMANENT DUOPOLY

- (DG, DG) is PSNE for exploration-disadvantaged MAB instances with a sufficiently small T_0

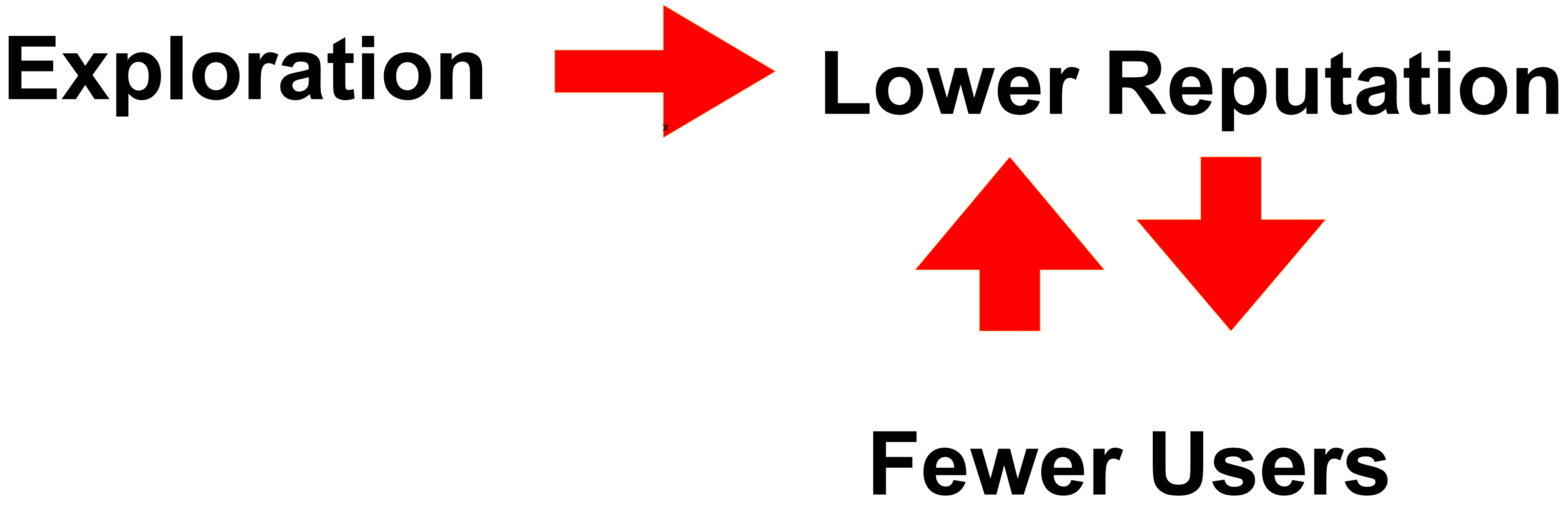


	$T_0 = 20$
TS vs DG	**0.29** ±0.03 EoG 55 (0)
TS vs DEG	**0.3** ±0.03 EoG 37 (0)
DG vs DEG	**0.62** ±0.03 EoG 410 (7)

- Effective End of Game (EoG) - the last round in our game when the firm choice between the agents at *t* and *t-1* differ.

PERMANENT DUOPOLY

- Note that the results reported before are for **T**=2000
- EoG values were much smaller than **T**!
- Evidence of ***death spiral effect:***
  - One firm attracts consumers at a slower rate than the other
  - Consequence: The other firm gets more consumers to learn from, which leads to faster learning, which leads to higher quality score



Exploration → **Lower Reputation**

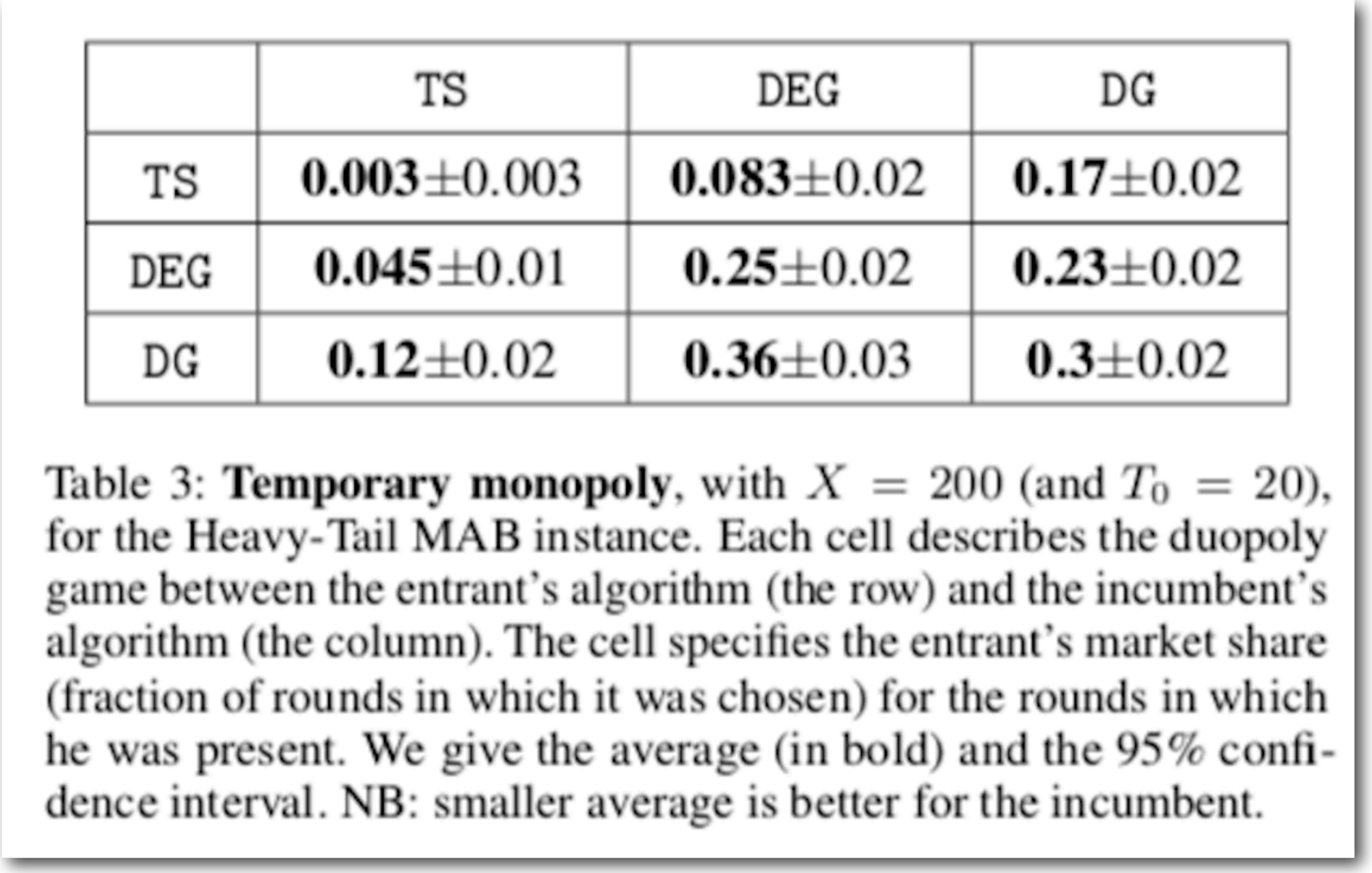
Fewer Users

EARLY ENTRY

- What happens if we let one firm enter the market early?

EARLY ENTRY

- What happens if we let one firm enter the market early?



	TS	DEG	DG
TS	**0.003**±0.003	**0.083**±0.02	**0.17**±0.02
DEG	**0.045**±0.01	**0.25**±0.02	**0.23**±0.02
DG	**0.12**±0.02	**0.36**±0.03	**0.3**±0.02

Table 3: **Temporary monopoly**, with $X = 200$ (and $T_0 = 20$), for the Heavy-Tail MAB instance. Each cell describes the duopoly game between the entrant's algorithm (the row) and the incumbent's algorithm (the column). The cell specifies the entrant's market share (fraction of rounds in which it was chosen) for the rounds in which he was present. We give the average (in bold) and the 95% confidence interval. NB: smaller average is better for the incumbent.

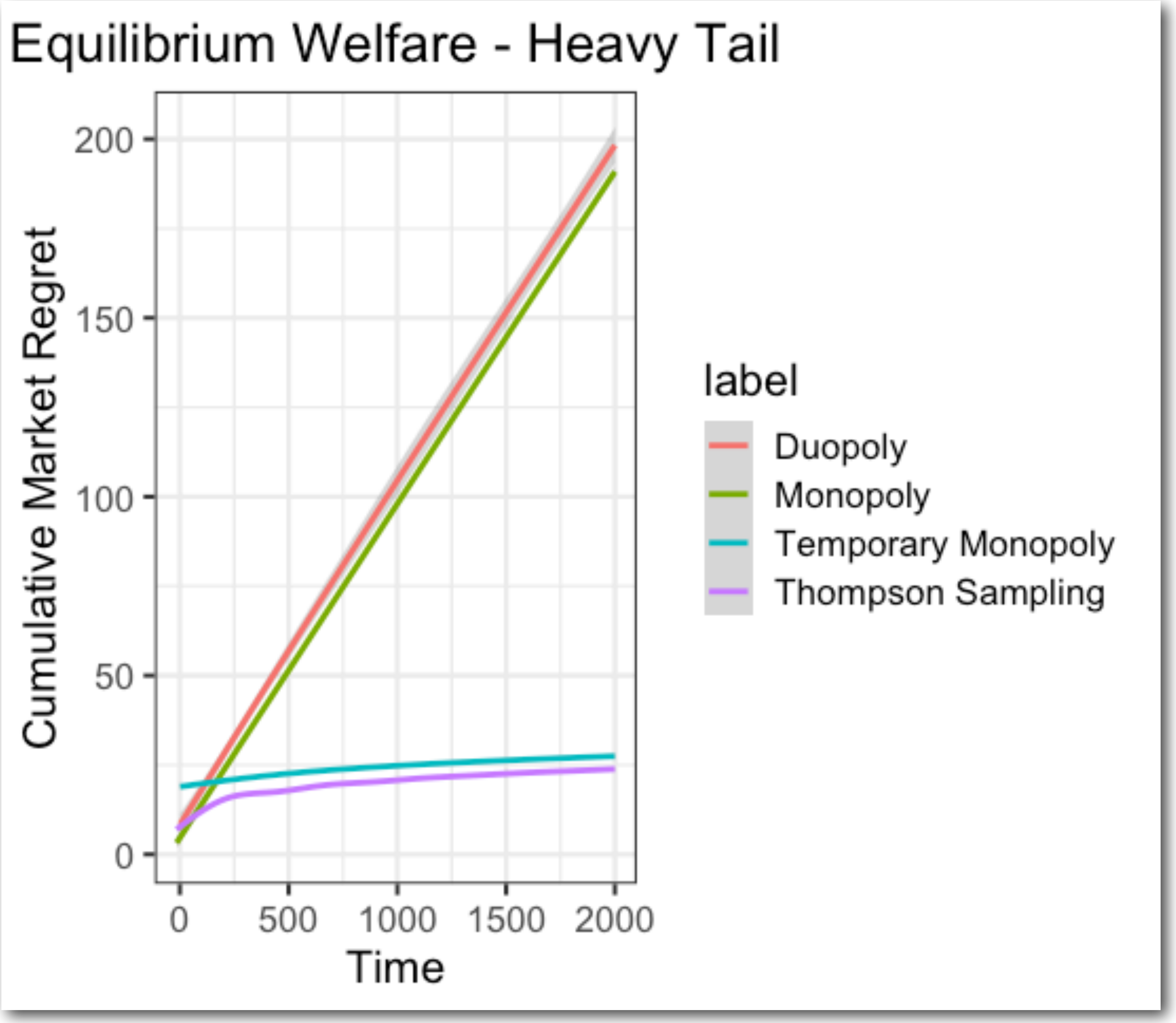
- Incumbent incentivized to play TS and entrant incentivized to play DG!

EARLY ENTRY

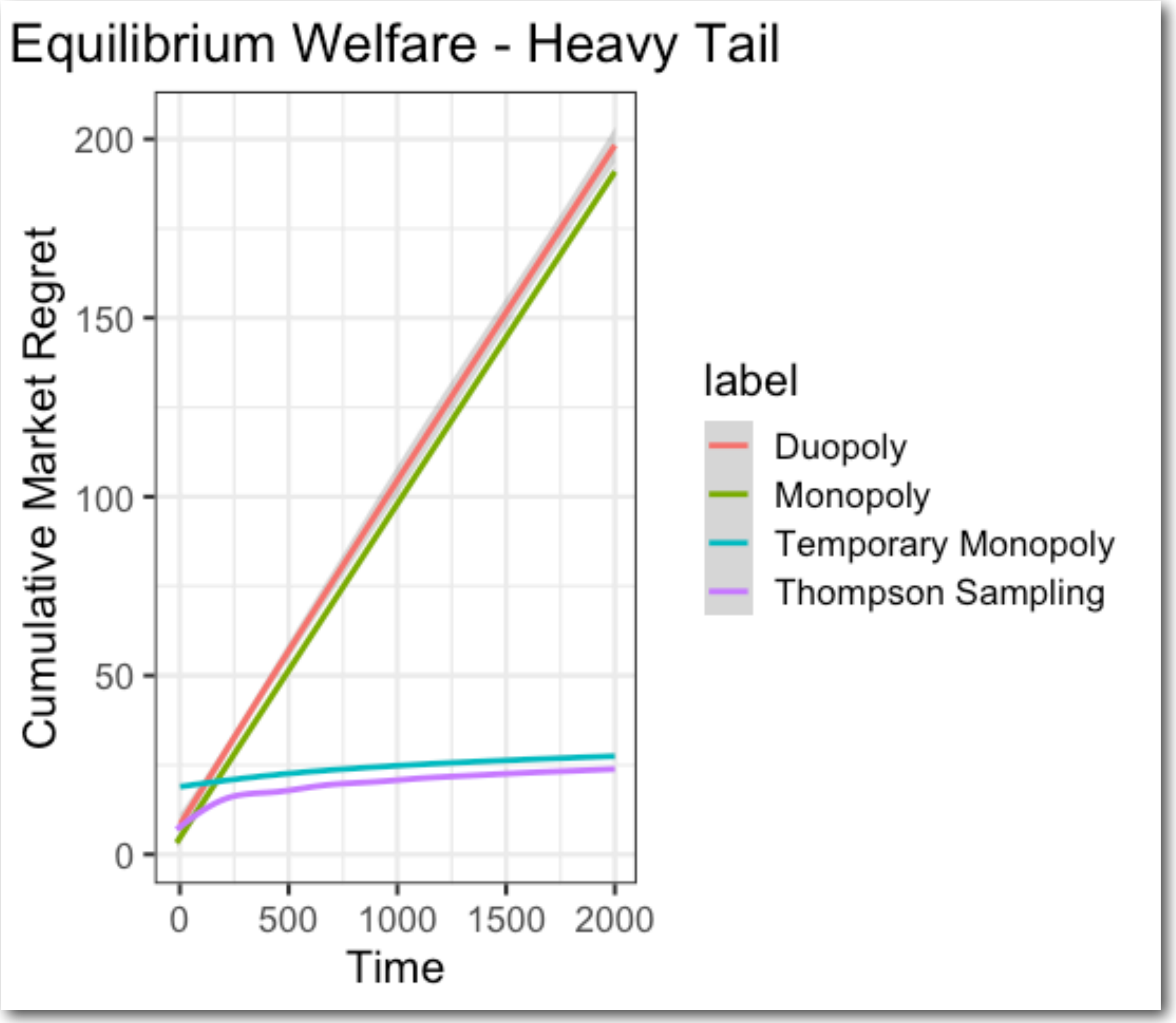
INTUITION

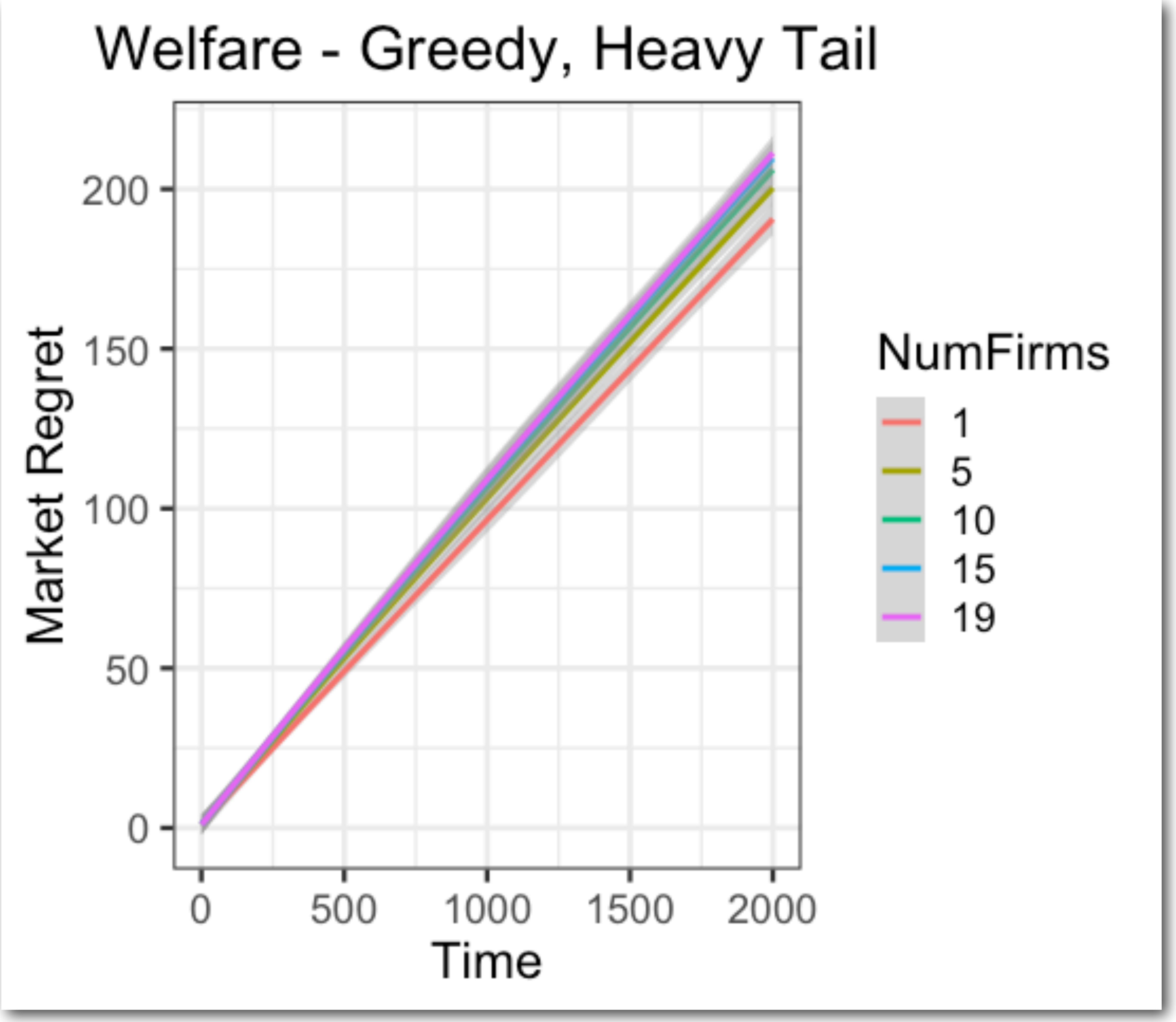
- Allows incumbent to not have to worry about the immediate reputation consequences of exploration!
- For sufficiently large **X**,
  - incumbent only faces the classic exploration-exploitation tradeoff
  - picks algorithms that are best at optimizing this tradeoff
  - recovers the reputation consequences of exploration

WELFARE


Equilibrium Welfare - Heavy Tail
Cumulative Market Regret
200
150
100
50
0
0
500
1000
1500
2000
Time
label
Duopoly
Monopoly
Temporary Monopoly
Thompson Sampling

WELFARE


Equilibrium Welfare - Heavy Tail
Cumulative Market Regret
0
50
100
150
200
0
500
1000
1500
2000
Time
label
Duopoly
Monopoly
Temporary Monopoly
Thompson Sampling

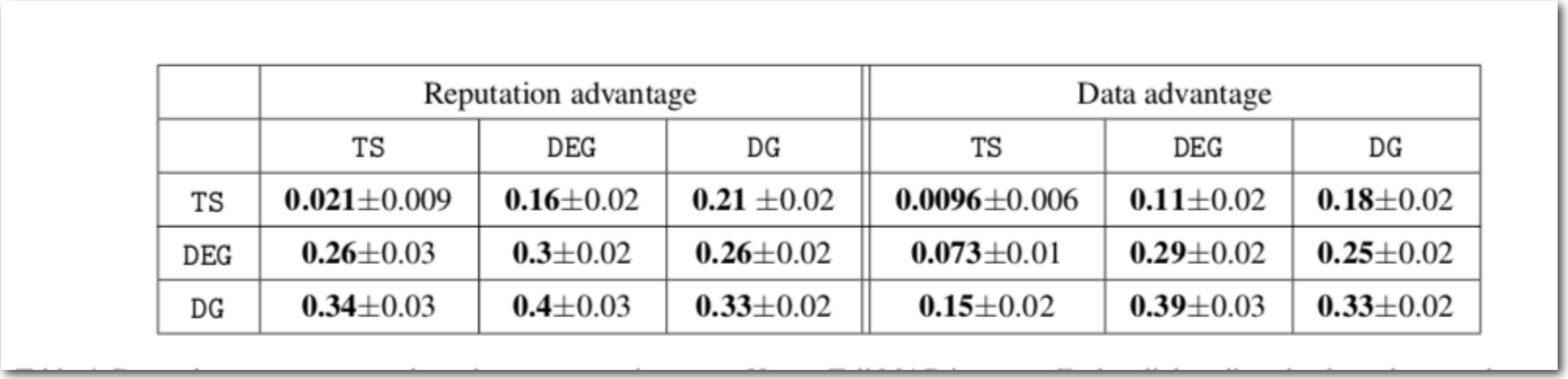

Welfare - Greedy, Heavy Tail
Market Regret
0
50
100
150
200
0
500
1000
1500
2000
Time
NumFirms
1
5
10
15
19

DATA AND REPUTATION AS BARRIERS TO ENTRY

- In the early entry case, the incumbent gets a substantial market share!
- Two advantages:
  - More definite (and possibly better) reputation (***reputation advantage***)
  - More data than entrant (***data advantage***)
- Natural question: Which advantage serves as larger barrier to entry (and when)?

DATA OR REPUTATION?

- Consider two experiments
  - Reputation advantage (reset incumbent posterior to fake prior upon entry)
  - Data advantage (reset incumbent reputation score upon entry)



	Reputation advantage			Data advantage		
	TS	DEG	DG	TS	DEG	DG
TS	**0.021**±0.009	**0.16**±0.02	**0.21** ±0.02	**0.0096**±0.006	**0.11**±0.02	**0.18**±0.02
DEG	**0.26**±0.03	**0.3**±0.02	**0.26**±0.02	**0.073**±0.01	**0.29**±0.02	**0.25**±0.02
DG	**0.34**±0.03	**0.4**±0.03	**0.33**±0.02	**0.15**±0.02	**0.39**±0.03	**0.33**±0.02

DATA AND REPUTATION AS BARRIERS TO ENTRY

TAKEAWAYS

- Retaining the data or reputation advantage alone retains large market share!
- Data advantage is larger ***when the incumbent commits to TS*** (otherwise, roughly the same)
- When thinking about data as a barrier to entry
  - Data quality (and not just quantity) matters
  - Riskier actions harder to gather data on in competition

CONCLUSION

- Considered a model of competition between learning algorithms
- “Better algorithms” not always better in competition due to the reputational consequences of exploration
- Data can serve as a barrier to entry in online platforms, especially when exploration has reputation costs