THE PERILS OF EXPLORATION UNDER COMPETITION: A COMPUTATIONAL MODELING APPROACH

GUY ARIDOR
COLUMBIA ECONOMICS

KEVIN LIU
COLUMBIA CS

ALEX SLIVKINS
MICROSOFT RESEARCH, NYC

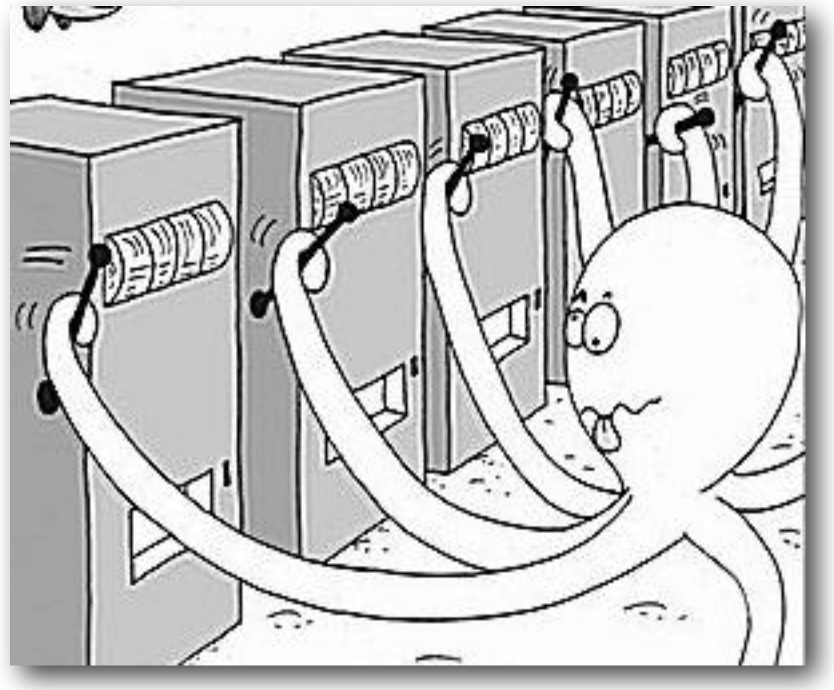
STEVEN WU
UNIVERSITY OF MINNESOTA CS

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

OUR SCOPE

- 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**
- Questions:
  - Does competition incentivize adoption of better algorithms?
  - What is the role that data and reputation play as barriers to entry?

(STOCHASTIC) MULTI-ARMED BANDITS

- In each period, select an action (“arm”) from a fixed set of arms, observe (random) reward for this arm, and nothing else
  - mean reward of each arm is fixed over time but not known
  - Goal: maximize cumulative reward over **T** periods.
- Captures exploration-exploitation tradeoff
  - **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*

OUR MODEL

- **Two** firms, both face the same bandit problem
  - **K arms: different ways to serve a user**
  - **Initially, each firm commits to a bandit algorithm**
  - Warm start: T_0 rounds before the competition starts
- In each round: new user arrives and chooses a firm,
  the firm chooses an "arm", the user receives a reward
  - **Reward is only observed by the chosen firm**
- Each firm's goal: maximize its (**expected) market share**
- **User's choice driven by "reputation" (average reward over sliding window)**

INNOVATION VS COMPETITION

MAB ALGORITHMS

- Innovation: Utilize the distinction between three classes of **MAB learning algorithms.**
  - **Greedy:** take actions with maximal mean reward, based on the current information. Use Dynamic Greedy (DG)
  - **Exploration-Separating**: exploration choices do not use the rewards observed so far. Use Dynamic Epsilon-Greedy (DEG)
  - **Adaptive Exploration:** Sway exploration choices towards more promising alternatives. Use Thompson Sampling (TS)
  - In isolation, **Adaptive Exploration > Exploration-Separating > Greedy**

- Competition: vary **timing of entry** and **number of firms** in the market

METHODOLOGY

- Study our model via numerical simulation
- 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)
- <u>Each experiment</u>: competition between bandit algorithms
    - Parameters: bandit algorithms, competition model, bandit instance

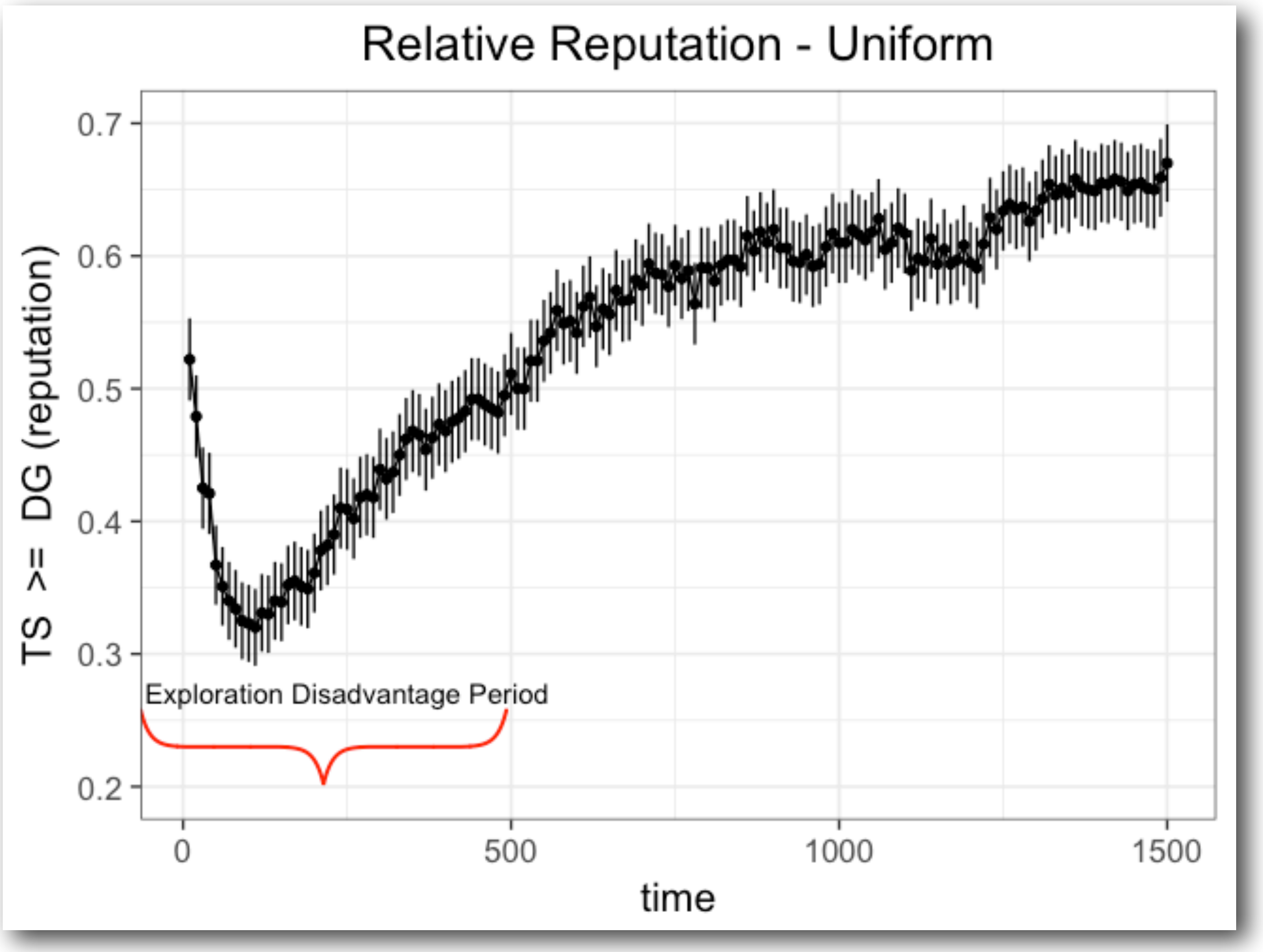
RELATED LITERATURE

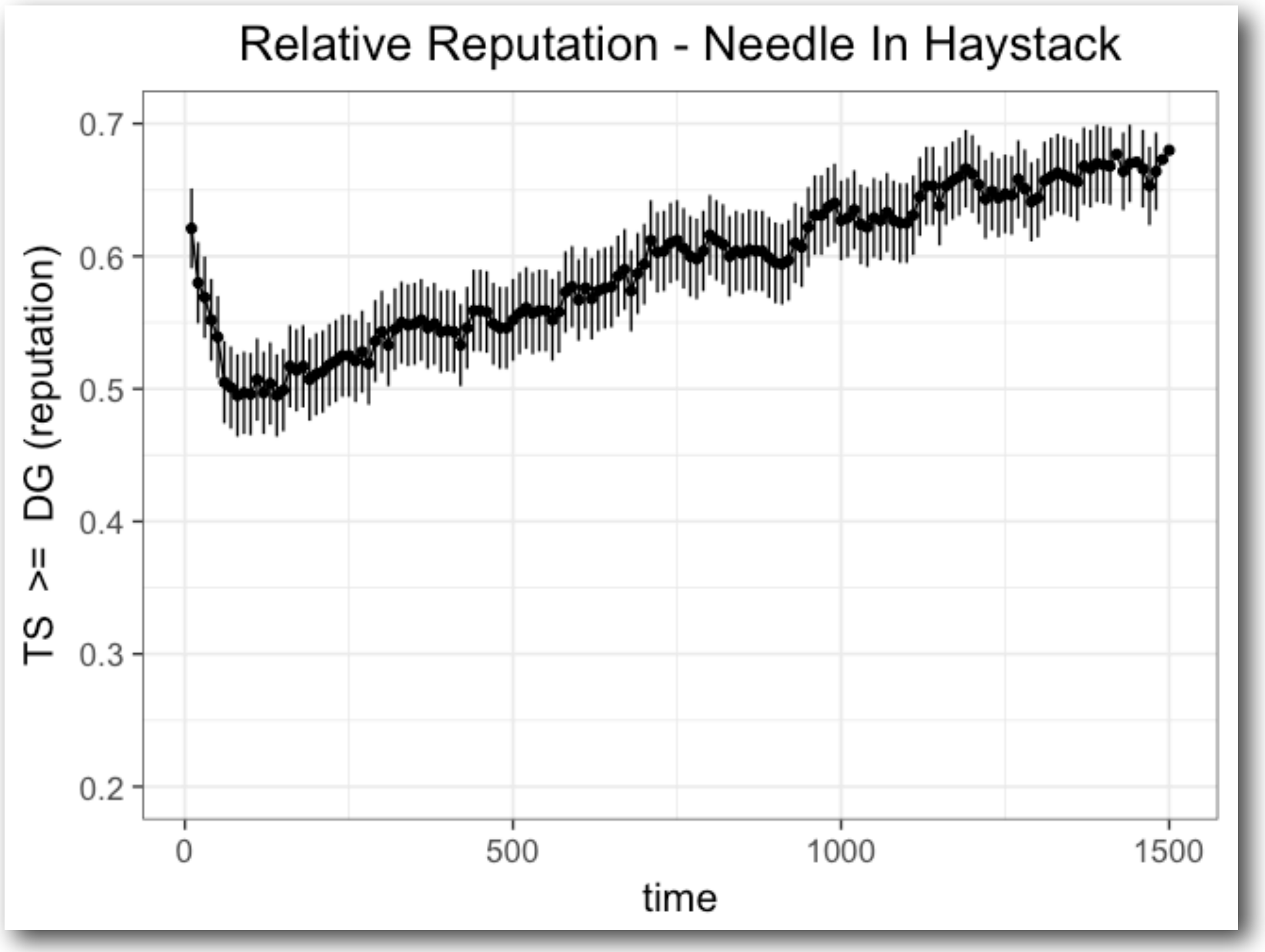
- Multi-armed bandits: well-studied model for exploration
  - Huge literature on bandit algorithms
- Bandit algorithms with incentives (large literature, diff scenarios):
  - “principal” runs a bandit algorithm,
    “agents” are bidders in an auction, users in a recommendation system, etc.
- Competition vs Innovation
  - In general: "Inverted-U relationship: Schumpter (1942), Aghion et.al (2005)
  - For exploration: (Mansour, Slivkins, Wu 2018)
    - different model: no “reputation”, competition varied via user response
    - Theory only, “asymptotic” results

STOPPED HERE

PERFORMANCE IN ISOLATION

- Mean reputation not a useful indicator of performance in competition
- Track ***relative reputation*** - at a given time t, fraction of simulations in which Alg 1 has a higher reputation score than Alg 2
- Purposeful exploration can lead to short-term ***reputation consequences***
    - When this occurs, call the instance ***exploration-disadvantaged***





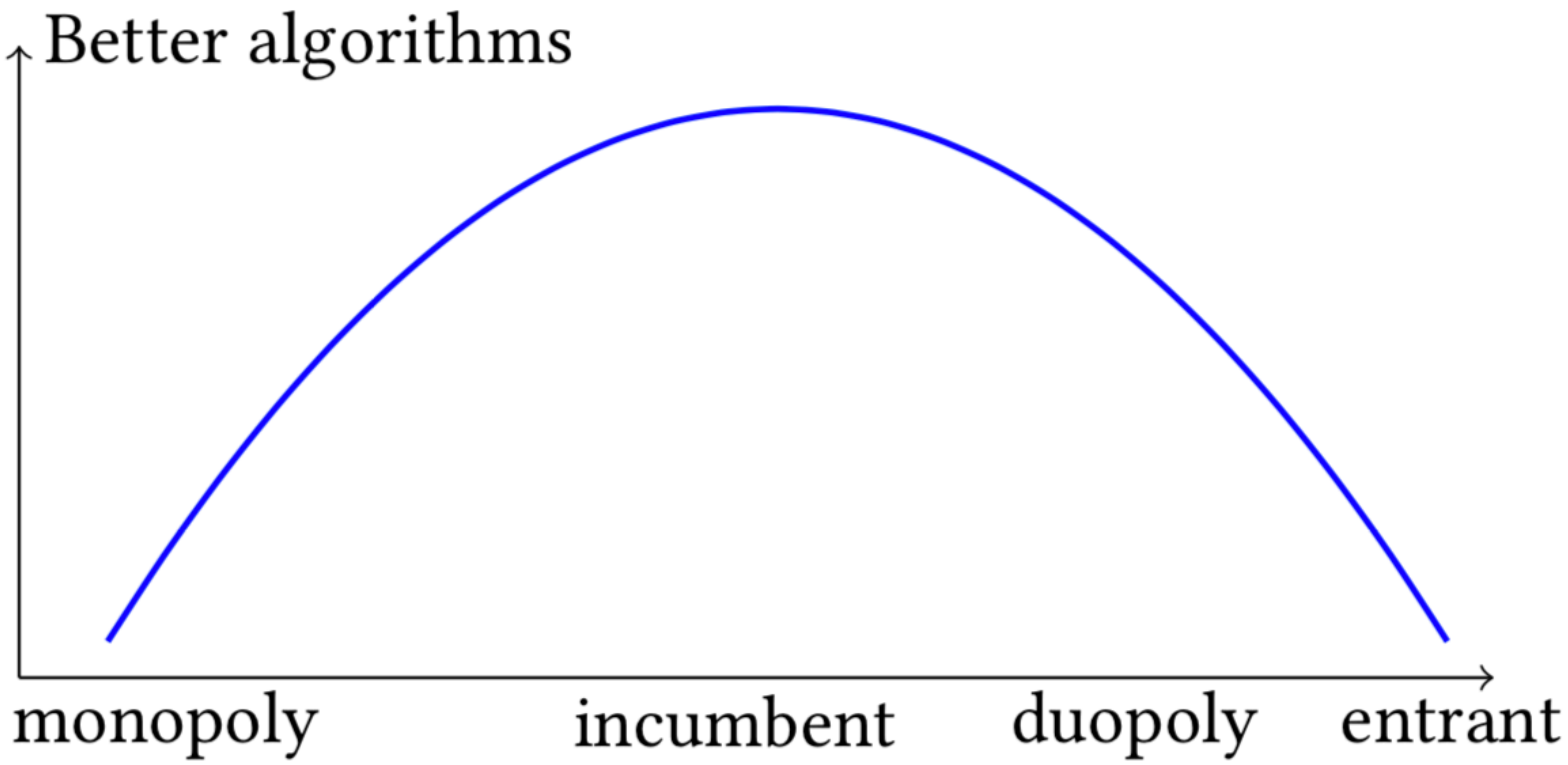
COMPETITION INTENSITY LEVELS

- We find the equilibrium strategies for four separate "competition intensity" levels:

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. 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:



DUOPOLY

EQUILIBRIUM

- Use simulation to compute expected market share (payoffs)
- For exploration-disadvantaged instances, (DG, DG) is a PSNE

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**

DUOPOLY

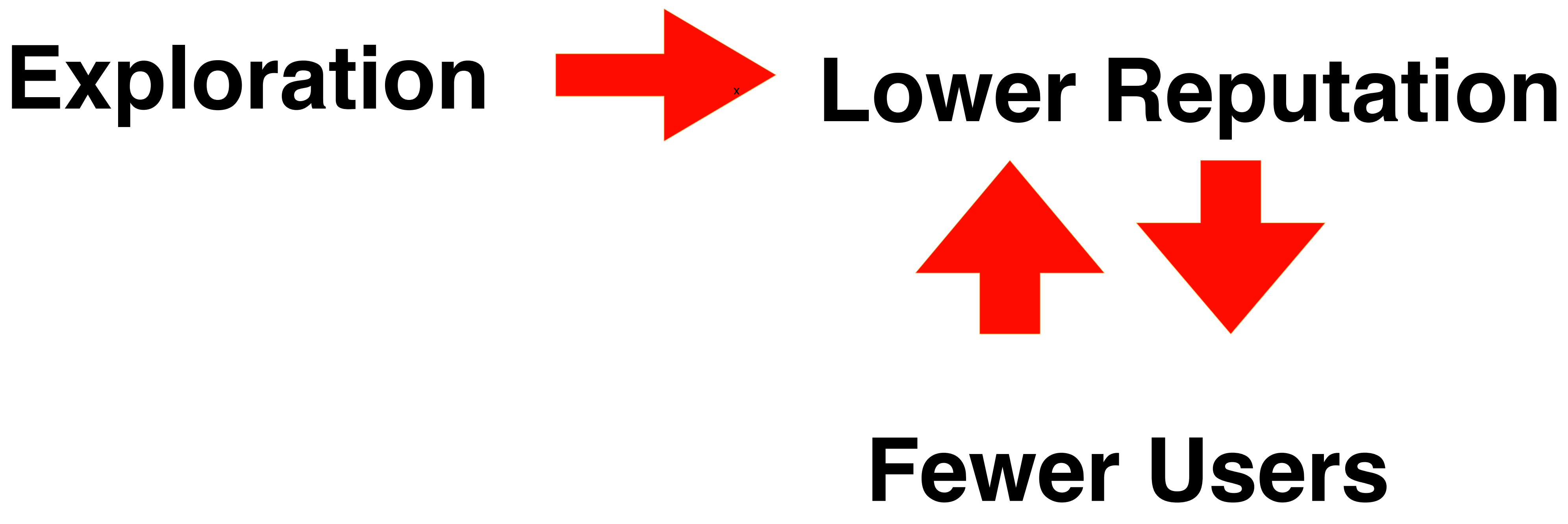
DEATH SPIRAL

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

	TS vs DG	TS vs DEG	DG vs DEG
EoG	55 (0)	37 (0)	410 (7)

Mean and Median EoG for Heavy-Tail Instance, $T = 2000$, $T_0 = 20$.

- Evidence of ***death spiral effect:***



EARLY ENTRY

EQUILIBRIUM

- Allow one firm to enter early and give it a ***temporary monopoly***
  - ***Incumbent*** (the early entrant) incentivized to play TS
  - ***Entrant*** (the late entrant) incentivized to play DG

	TS	DEG	DG
TS	**0.003**$\pm$0.003	**0.083**$\pm$0.02	**0.17**$\pm$0.02
DEG	**0.045**$\pm$0.01	**0.25**$\pm$0.02	**0.23**$\pm$0.02
DG	**0.12**$\pm$0.02	**0.36**$\pm$0.03	**0.3**$\pm$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.

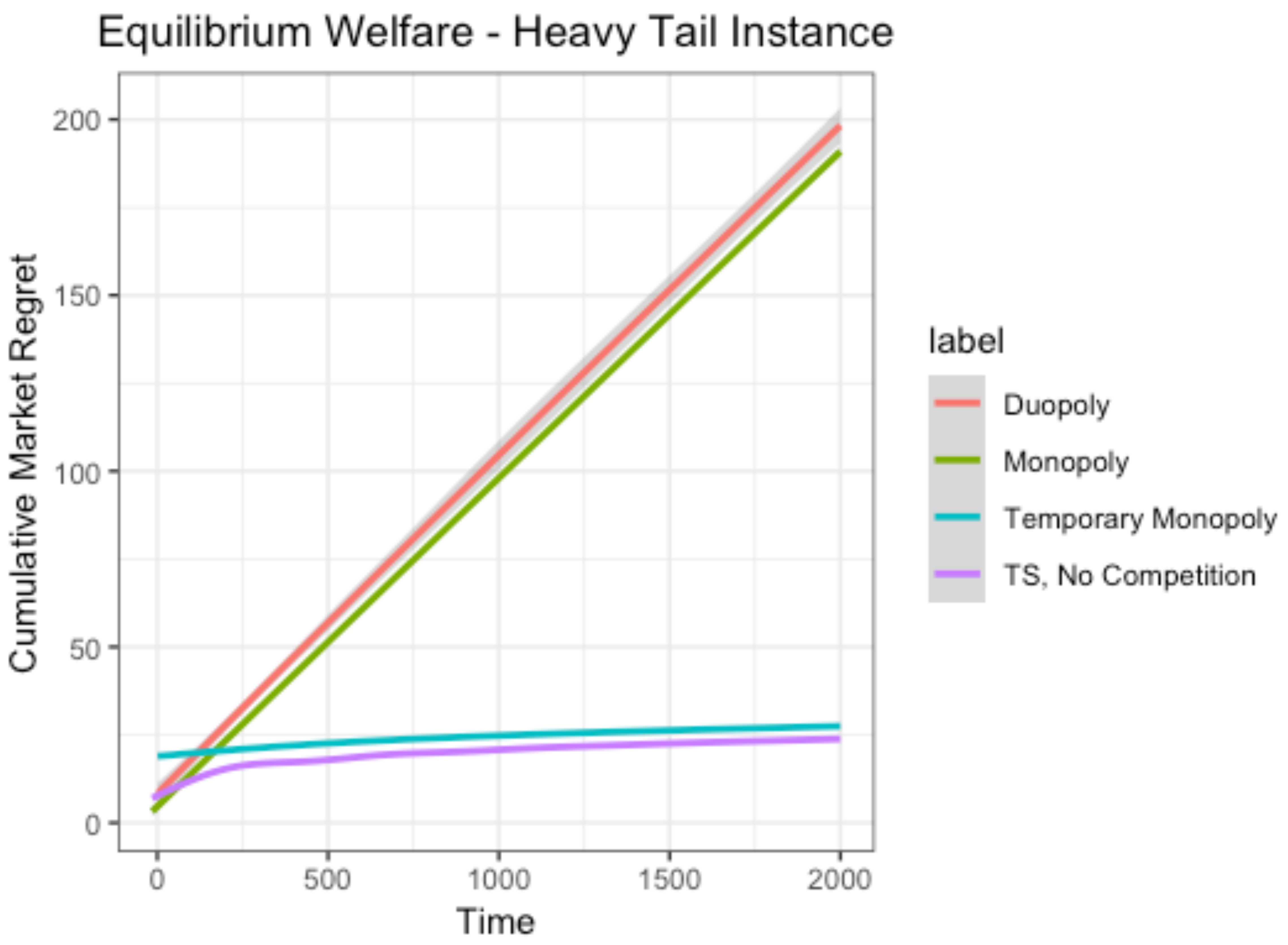
EARLY ENTRY

INTUITION

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

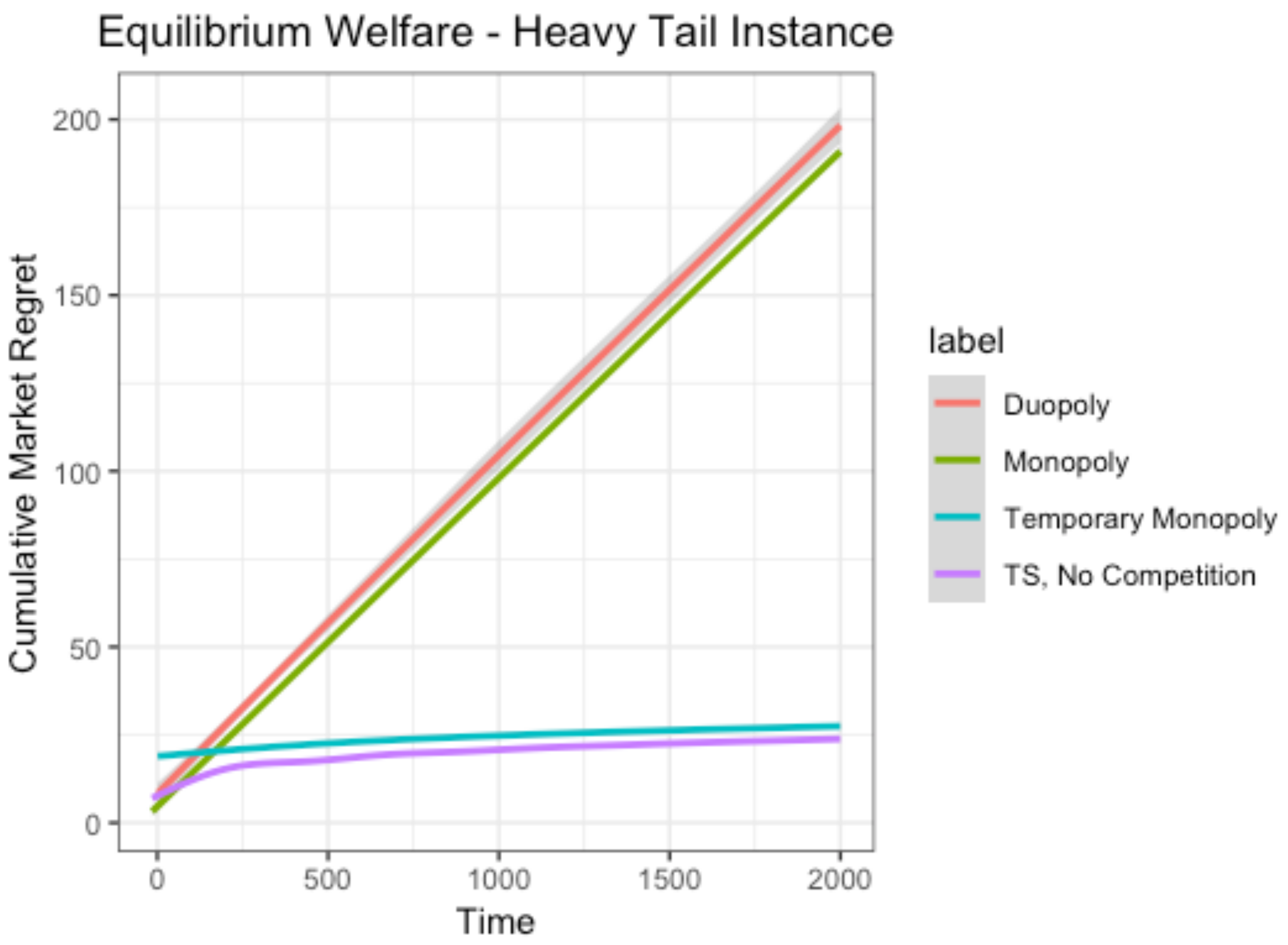
WELFARE

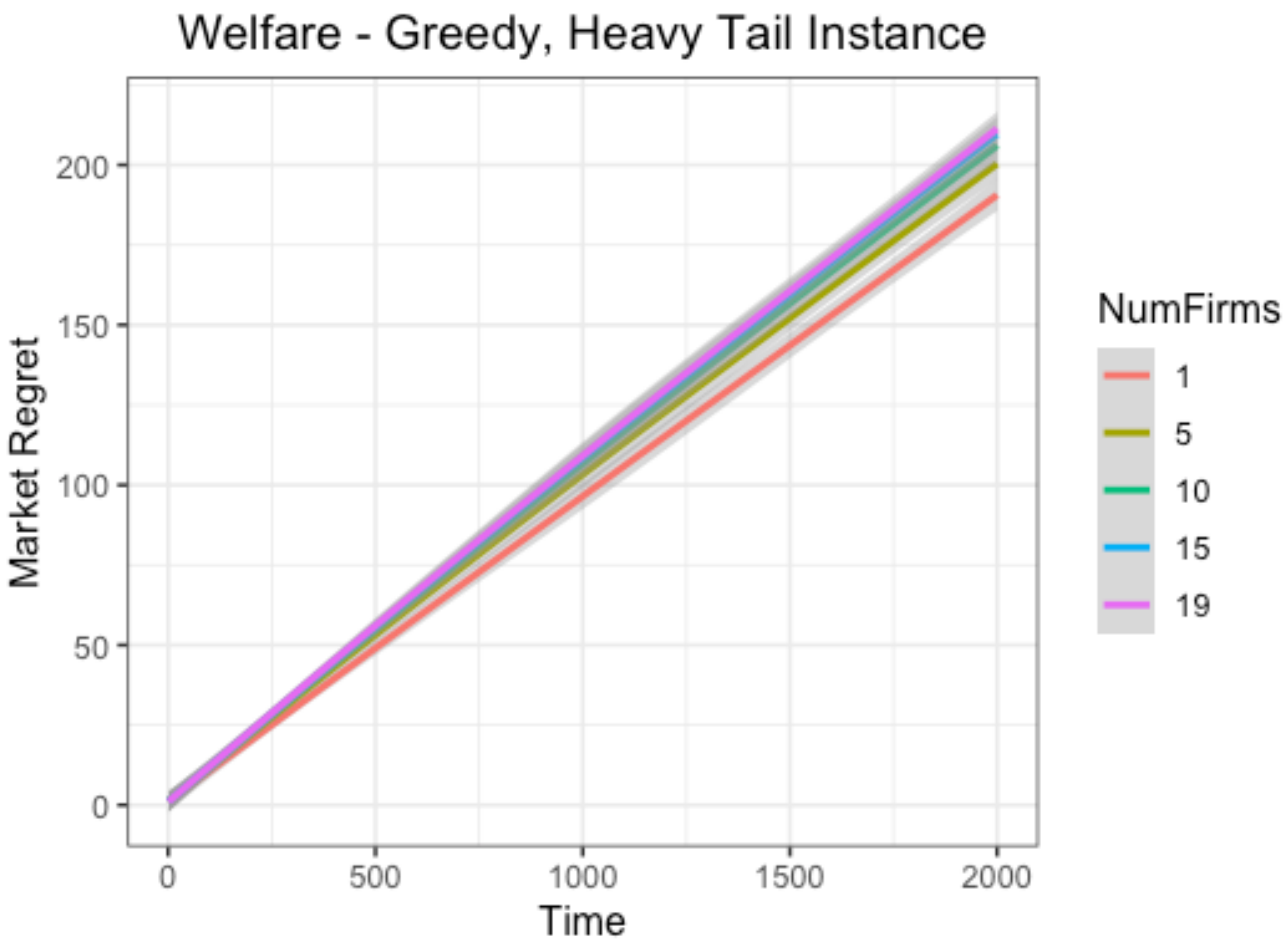
- Welfare measure = total reward accrued by all consumers
- As in MAB literature, focus on total *regret*



WELFARE

- Welfare measure = total reward accrued by all consumers
- As in MAB literature, focus on total *regret*





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 (only)			Data advantage (only)		
	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

User share of row player (entrant) after T = 2000

DATA AND REPUTATION AS BARRIERS TO ENTRY

TAKEAWAYS

- Retaining the data or reputation advantage alone leads to large market share!
- Data advantage is larger ***when the incumbent commits to TS***
- When thinking about data as a barrier to entry
  - Data quality (and not just quantity) matters
  - Depends on whether exploration has reputational consequences

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 reputational consequences