COMPETING BANDITS: THE PERILS OF EXPLORATION UNDER COMPETITION

GUY ARIDOR
COLUMBIA

YISHAY MANSOUR
TEL AVIV U / GOOGLE

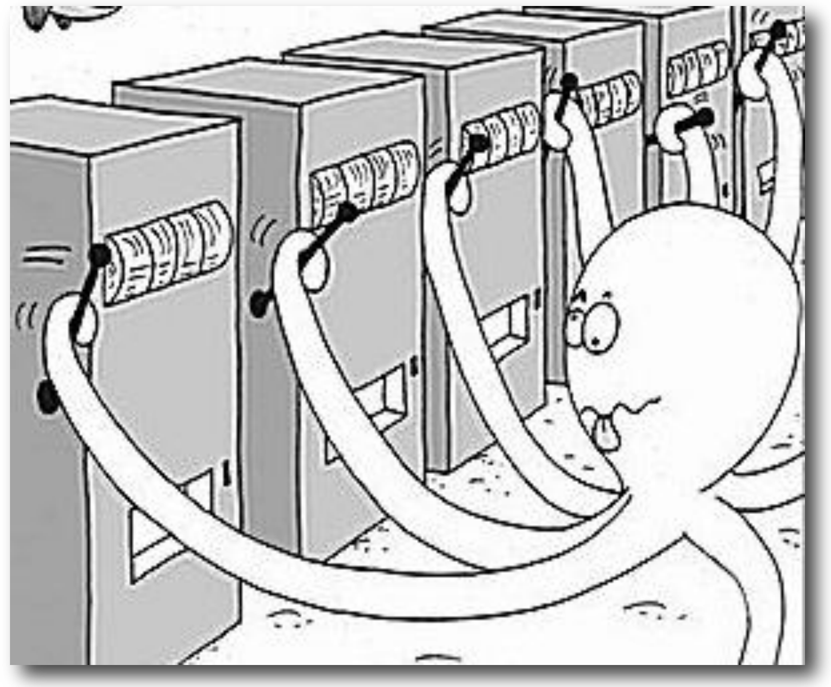
ALEX SLIVKINS
MICROSOFT RESEARCH, NYC

STEVEN WU
CARNEGIE MELLON

2020 CONFERENCE ON DIGITAL EXPERIMENTATION

MOTIVATION

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



- Platforms do not exist in isolation – ***compete*** against other platforms for users

- <u>This paper</u>: Firms ***compete*** for users and ***learn*** from the data generated by them

OUR SCOPE

- Study the tradeoff between ***exploration*** and ***competition:***
  1. Need to ***incentivize*** users to choose me over competition ***today***
  2. Need to ***explore*** to gain information to have a better product ***tomorrow***

- <u>Questions we study:</u>
  - When does competition incentivize the adoption of better exploration algorithms?
  - Can data play a similar role as traditional "network effects" in online markets?

(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(S)

- **Model Components:**
  - Two firms face identical MAB problems
  - Each firm's objective function: maximize its (**expected) market share**
  - Commit to MAB algorithms at **t = 0**
- **In each round:**
  - A single new user arrives and selects one firm
  - Chosen firm pulls an arm and user receives a reward
- **Reward is only observed by chosen firm.**
- **User choice driven by:**
  - Average reward over sliding window - Frequentist
  - Bayesian expected reward (BER) - Bayesian

INNOVATION

- Primary question: What is the equilibrium of the "competition" game?
    - Bayesian model: set of "monotone" MAB algorithms
    - Frequentist model: representative algorithm from three classes of MAB algorithms

- Utilize the distinction between three classes of MAB learning algorithms.

    - **Exploit Only** : pick arm with maximum mean reward based on current information.
    - **Exploration-Separating**: exploration does not depend on previous observations.
    - **Adaptive Exploration**: "smart" exploration based on previous observations

- Interpret the adoption of "better" algorithms as "innovation"

COMPETITION

- Vary competition by different choice models and timing of entry
  - Different choice models:

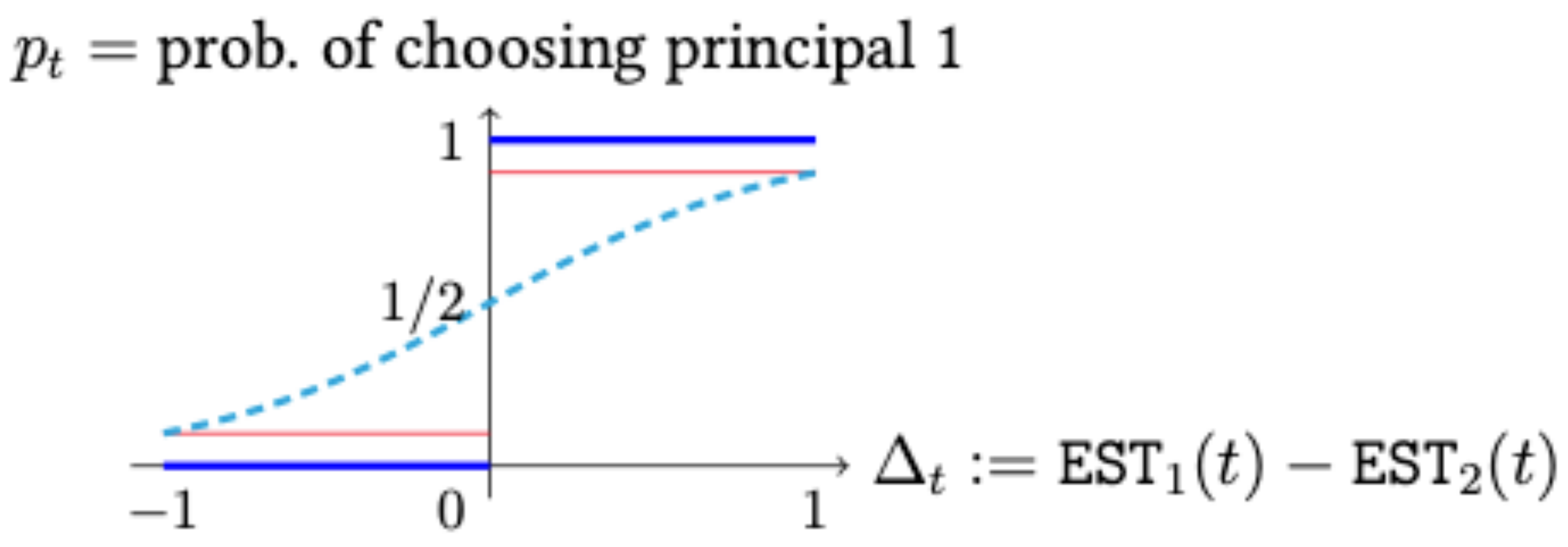
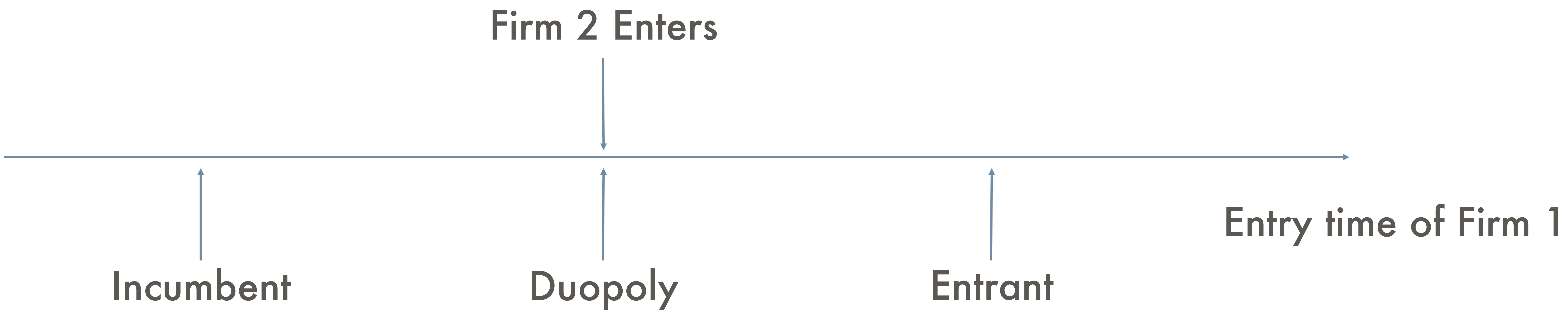


Figure 1: The three models for $f_{\texttt{resp}}$: `HardMax` is thick blue, `HardMax&Random` is red, and `SoftMax` is dashed.

  - Timing of entry:



EQUILIBRIUM UNDER “STARK” COMPETITION

- In both models, unique NE is for both firms to play the greedy algorithm
  - In Bayesian model: deviation at any time t leads to lower BER relative to greedy
  - In Frequentist model: deviation at any time t leads to ***death spiral effect***

Exploration → **Lower Reputation**

↑ ↓

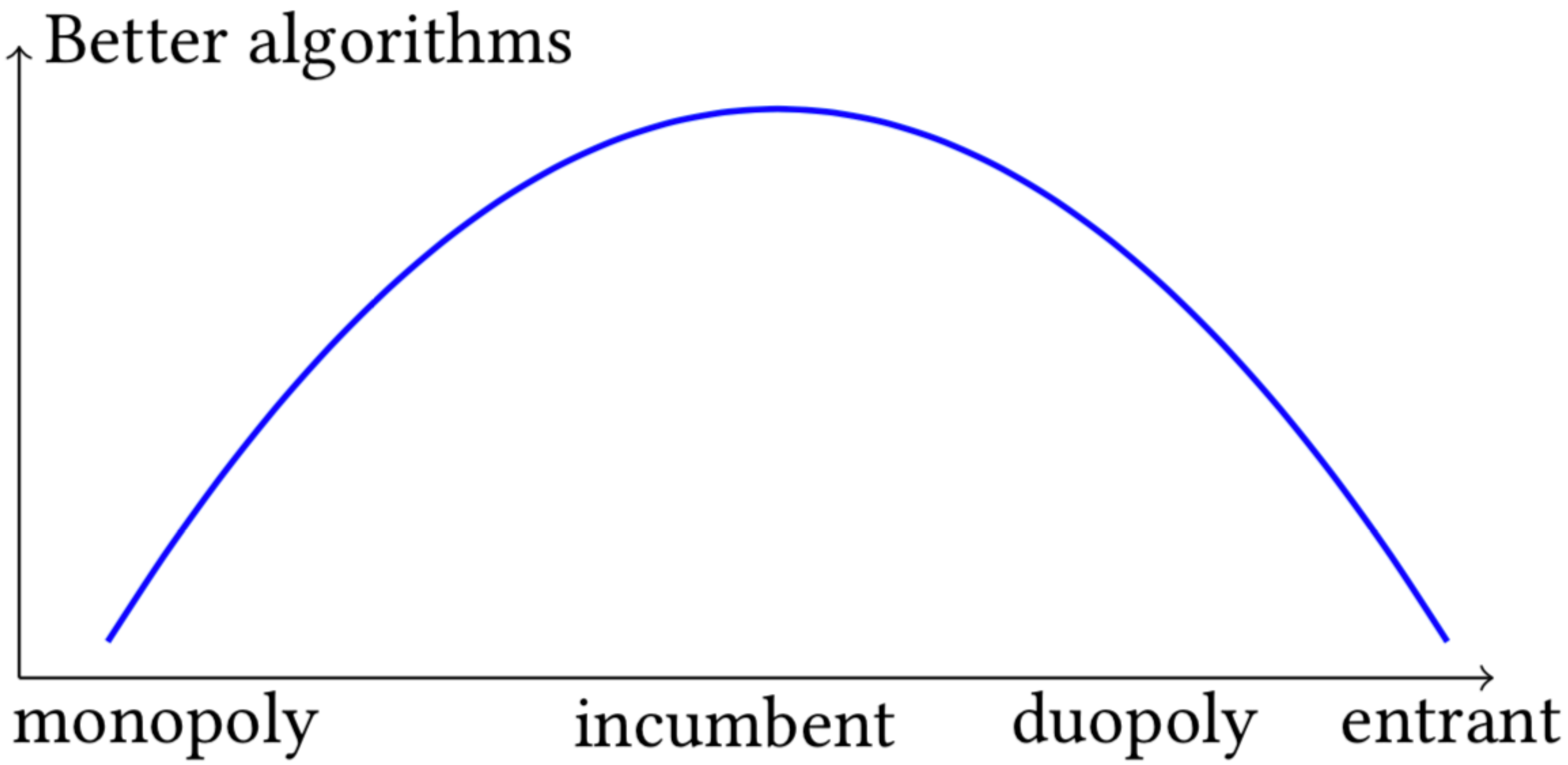
Fewer Users

- Key intuition: Reputational consequences of exploration disincentivize adoption of “better” algorithms

EQUILIBRIUM UNDER "WEAKENED" COMPETITION

- Equilibrium under different choice models:
  - With "random" users: better algorithms (sometimes) chosen in equilibrium
  - With "smooth" selection probabilities: better algorithms chosen in equilibrium
- Equilibrium under asymmetric timing of entry:
  - Incumbent commits to better algorithms, entrant commits to greedy algorithm
  - Incumbent garners vast majority of market share
- Higher consumer welfare in equilibrium (in both scenarios)
- <u>Key intuition</u>: Weakening the reputational consequences of exploration incentivizes adoption of better algorithms

SUMMARIZING MAIN RESULTS: INNOVATION VS COMPETITION



Stylized "inverted-U" relationship between competition and innovation

SUMMARIZING MAIN RESULTS: ROLE OF DATA

- Learning algorithms in isolation ≠ learning algorithms in competition
  - Data is an endogenous object determined by competition dynamics

- Small amount of data advantage amplified by competition dynamics
  - Similar to "data feedback loops" hypothesized in policy debates
  - Characterized and amplified not just by quantity, but also higher quality of data gathered by better learning algorithms