Crime Detection via Crowdsourcing

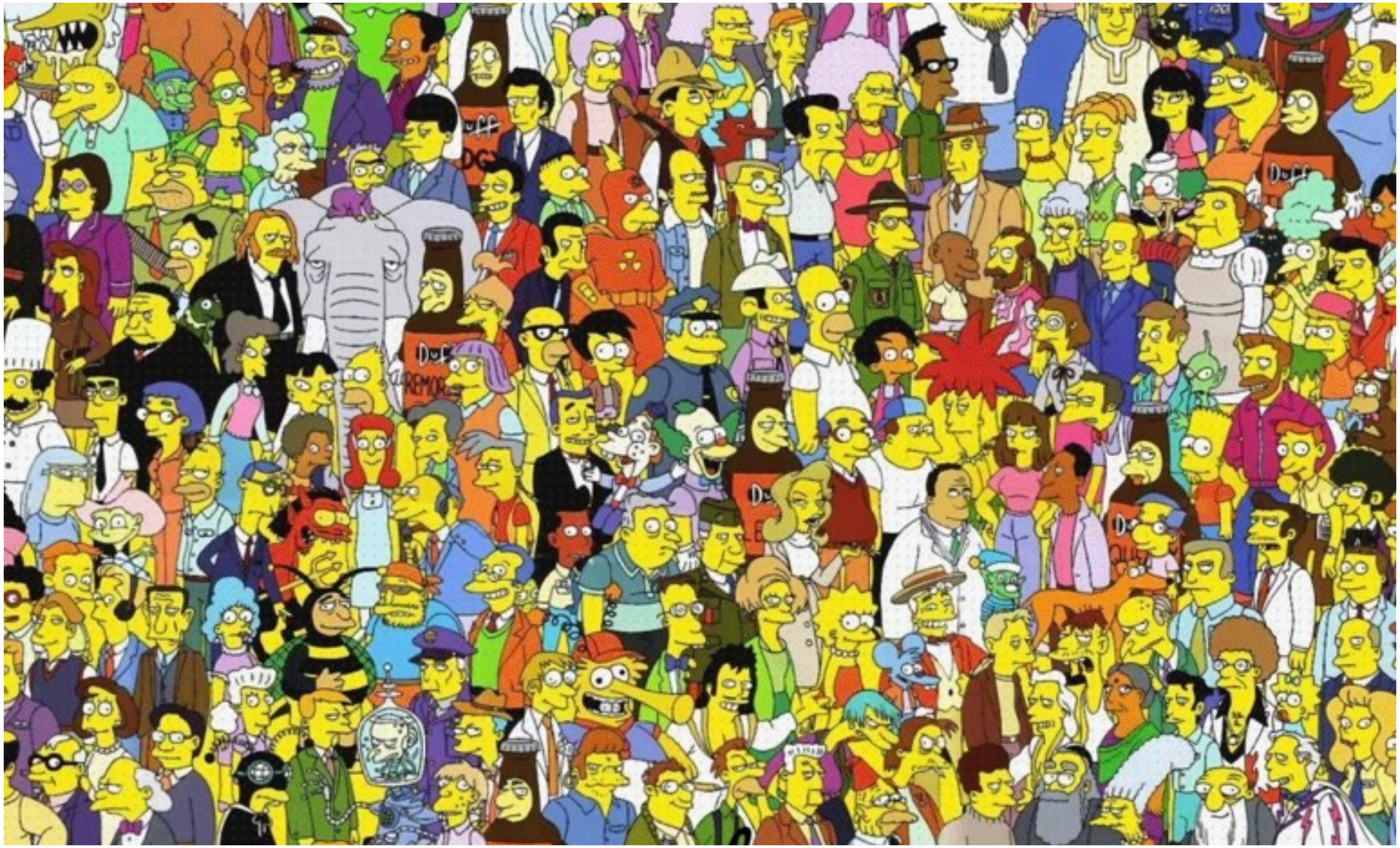
Daniel Pimentel & Claudia Solís

University of Wisconsin-Madison, USA

We want to catch criminals

Criminals leave evidence

Detective follows clues



Criminal could be anyone!

Needle in a haystack problem

Crowdsourcing

Obtain information by enlisting a lot of people

Crowdsourcing is an old idea

FBI asks community for information

Overwhelming task!

Let’s use Robots!

Let's use ~~Robots~~!

Next best thing: Computers!

Crowdsourcing is powerful!

Bird tracking

Crowdsourcing is powerful!

Finding lost aircrafts

Crowdsourcing is powerful!

Text Classification

So let's use Computers!

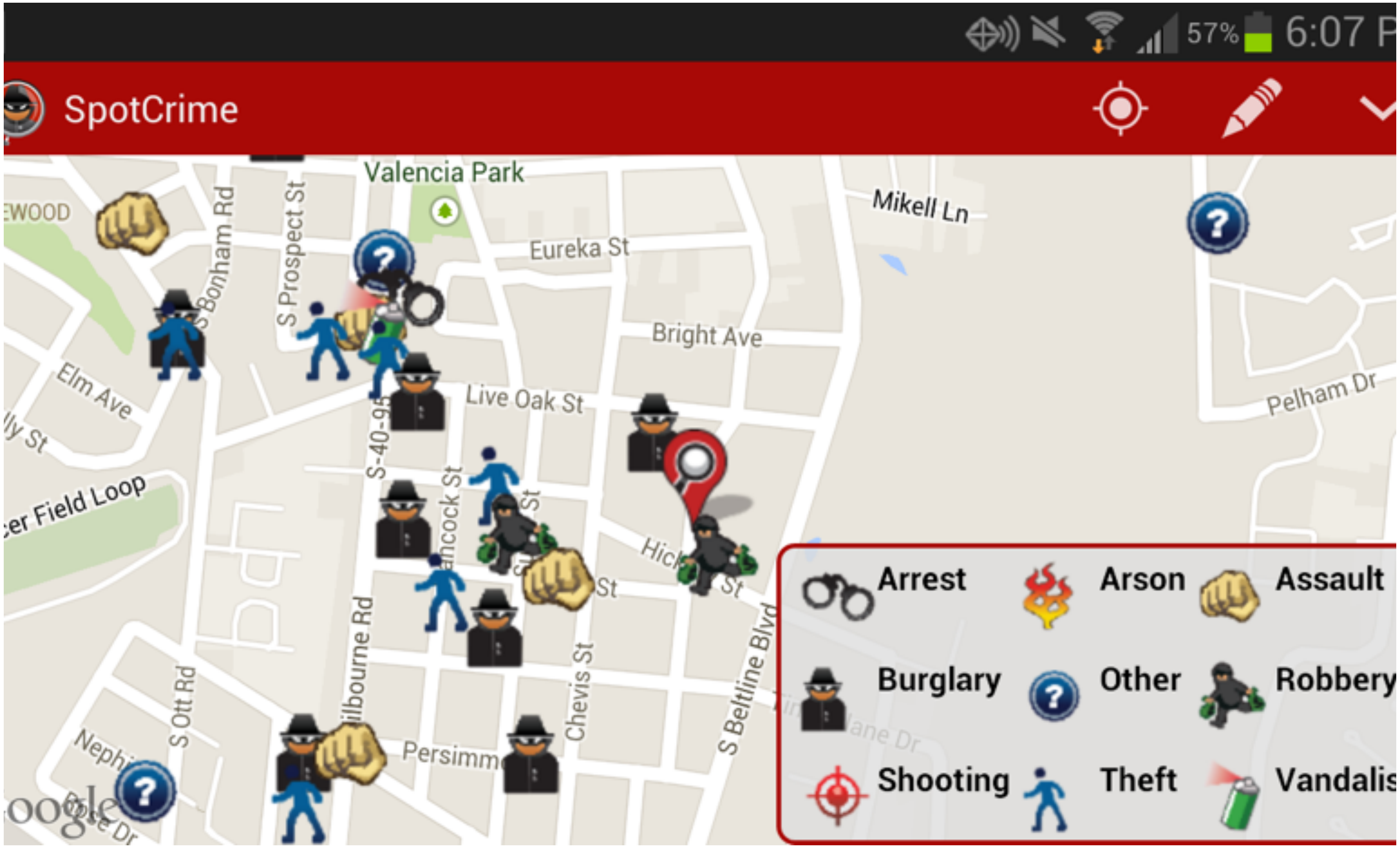
Conventional Criminal Tracking System

Conventional Criminal Tracking System

With Crowdsourcing

This works!



57%
6:07
SpotCrime
Valencia Park
Mikell Ln
Eureka St
Bright Ave
Live Oak St
Pelham Dr
S Bonham Rd
S Prospect St
Elm Ave
S-40-95
S Ott Rd
S Beltline Blvd
Chevis St
Arrest
Arson
Assault
Burglary
Other
Robbery
Shooting
Theft

This works?




Conventional Criminal Tracking System

With Crowdsourcing

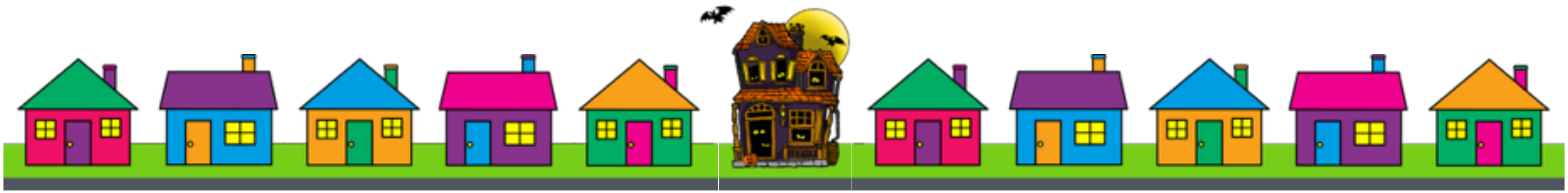
   

Let’s turn things around

New Criminal Tracking Model

Let’s change the rules of the game!



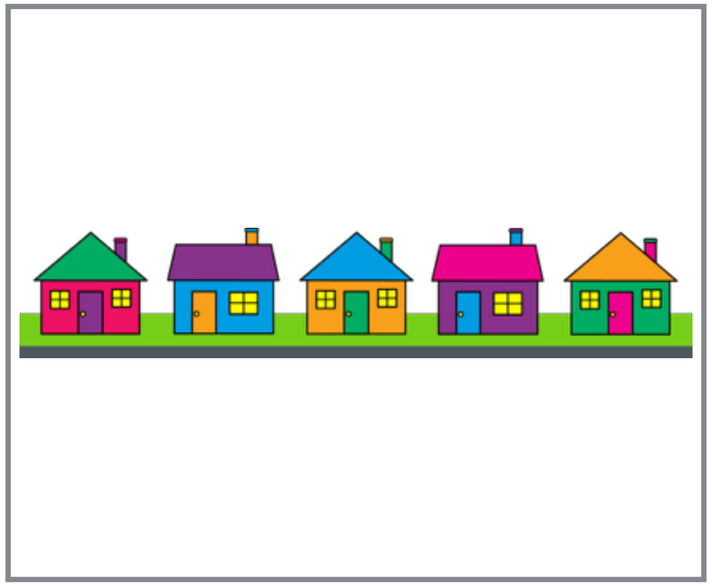
In Essence:

Classify Houses

In Essence:

Classify Houses

Could this ever work?



- 1 Criminal
- $n+1$ Houses: $h_1, h_2, \ldots, h_n, h_*$
- m Tips

Key Ingredients

New Criminal Tracking Model

- One man, one vote.
- Votes go to:
  - h_* with probability p_*,
  - h_j with probability p_j.
  - Assume $p_1 \geq p_2 \geq \cdots p_n$.
- Detective investigates $\hat{h}$.

Rules of the Game

New Criminal Tracking Model

Intuitively

- p_1 is the level of *prejudice*.
- p_* is the level of *accuracy*.

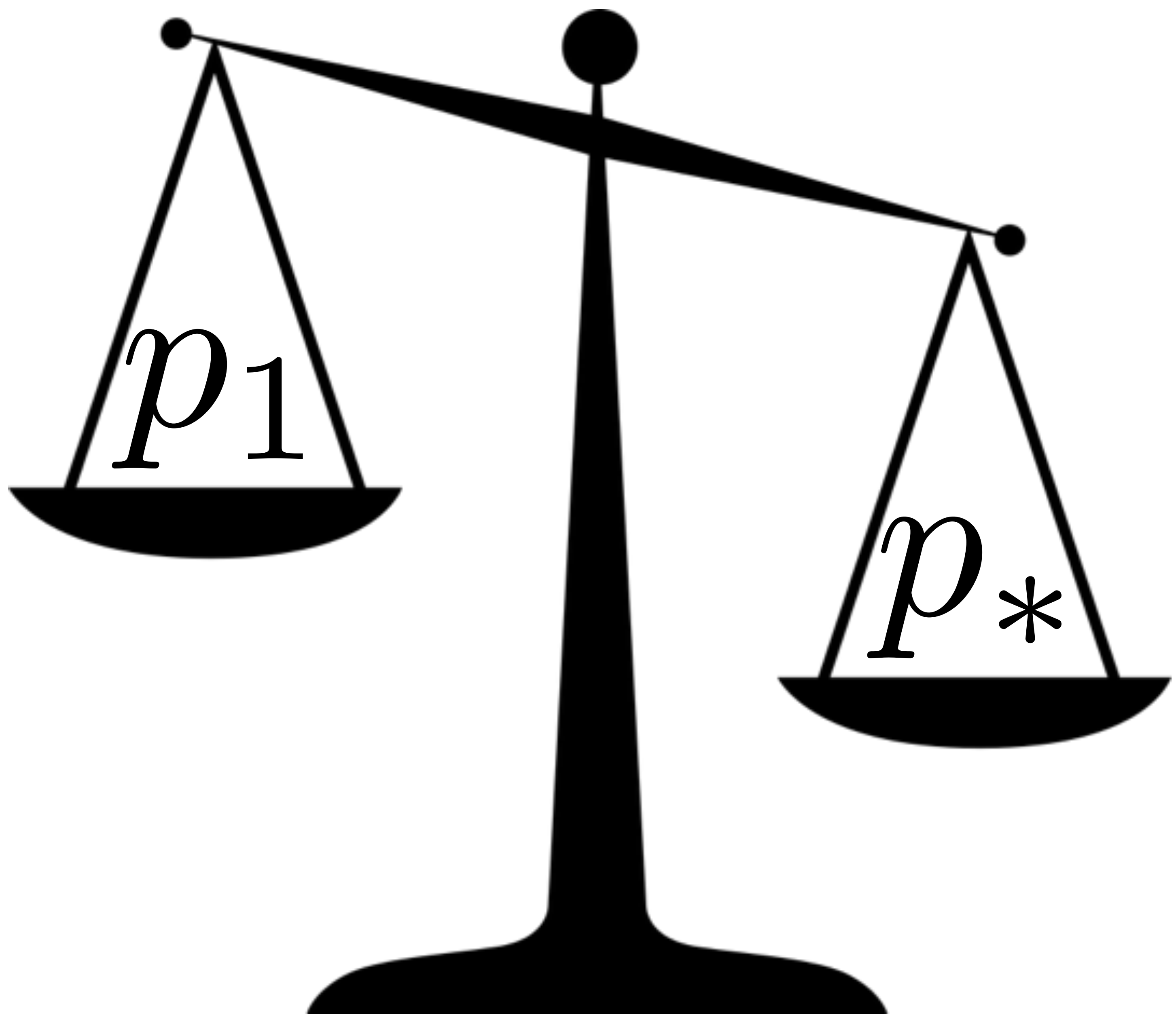
Theorem. Let $\epsilon > 0$ be given and suppose

$$p_* \;\geq\; p_1 + \sqrt{\frac{2}{m}\log\left(\frac{n}{\epsilon}\right)}.$$

Then $\hat{h} = h_*$ with probability at least $1 - \epsilon$.

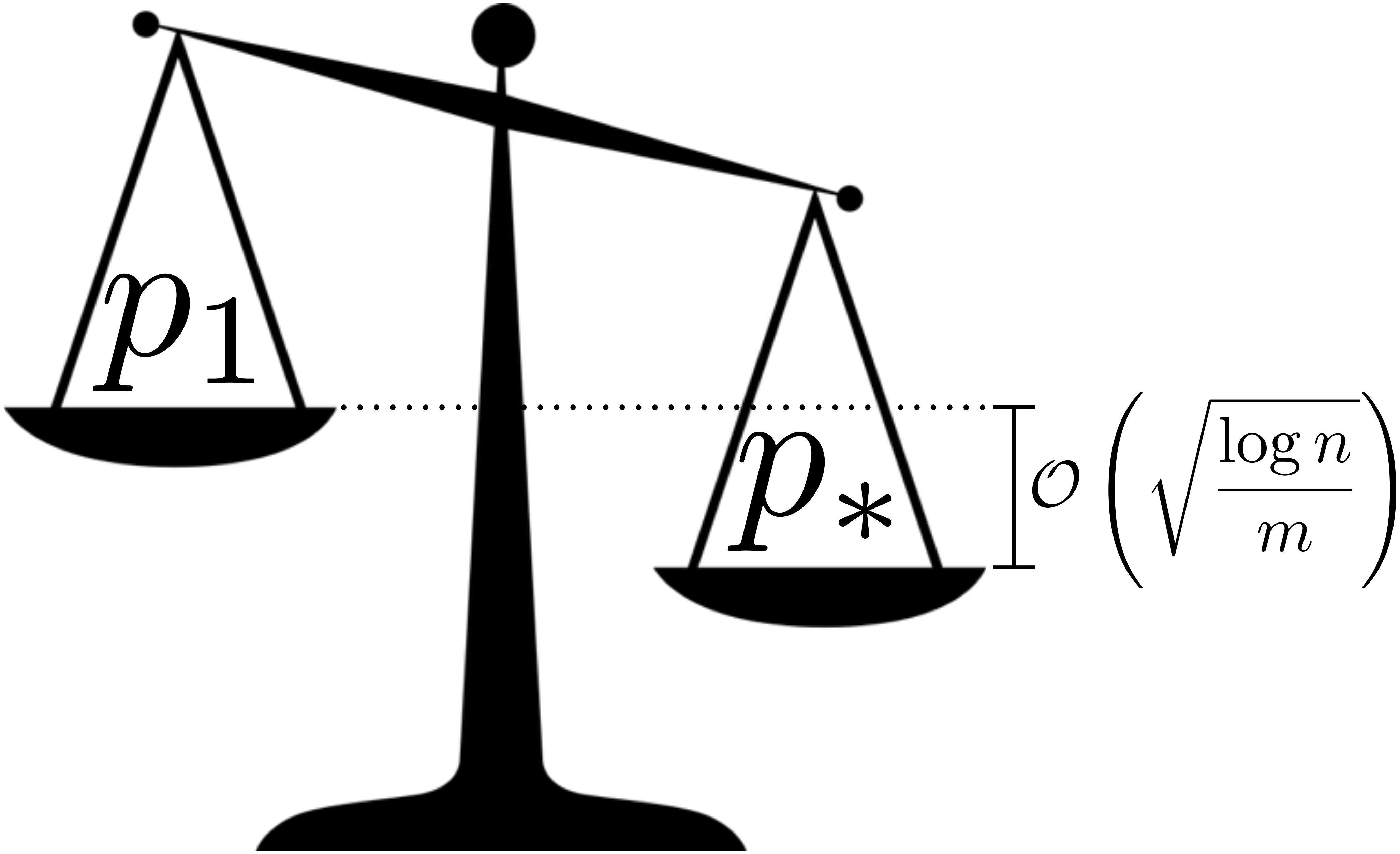
Main Result

New Criminal Tracking Model



In words

Main result matches intuition



In words

Main result matches intuition

$$
\begin{aligned}
\mathsf{P}(\hat{h} \neq h_\star) &= \mathsf{P}\Big(\bigcup_{j=1}^{n} \{N_\star \leq N_j\}\Big) \\
&\leq \sum_{j=1}^{n} \mathsf{P}(N_\star \leq N_j) && \longleftarrow \text{Union Bound} \\
&\leq \sum_{j=1}^{n} \mathsf{P}(N_\star \leq N_1) && \longleftarrow \text{Assumption} \\
&\leq e^{-\frac{m}{2}(p_\star - p_1)^2} && \longleftarrow \text{Hoeffding's} \\
&\leq \epsilon.
\end{aligned}
$$

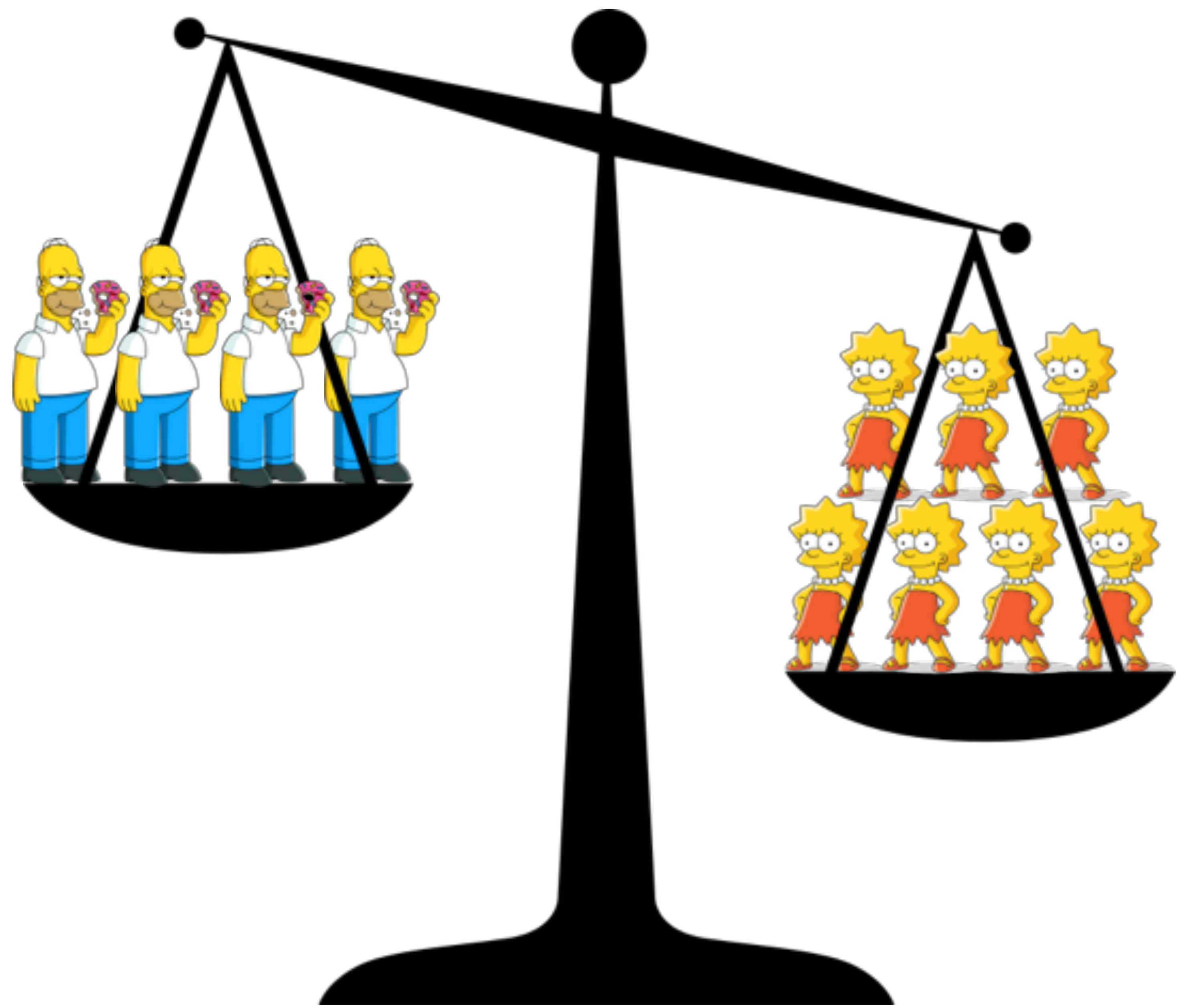
Key Idea of Proof

Geography & Bias

Model Extensions

Geography & Bias

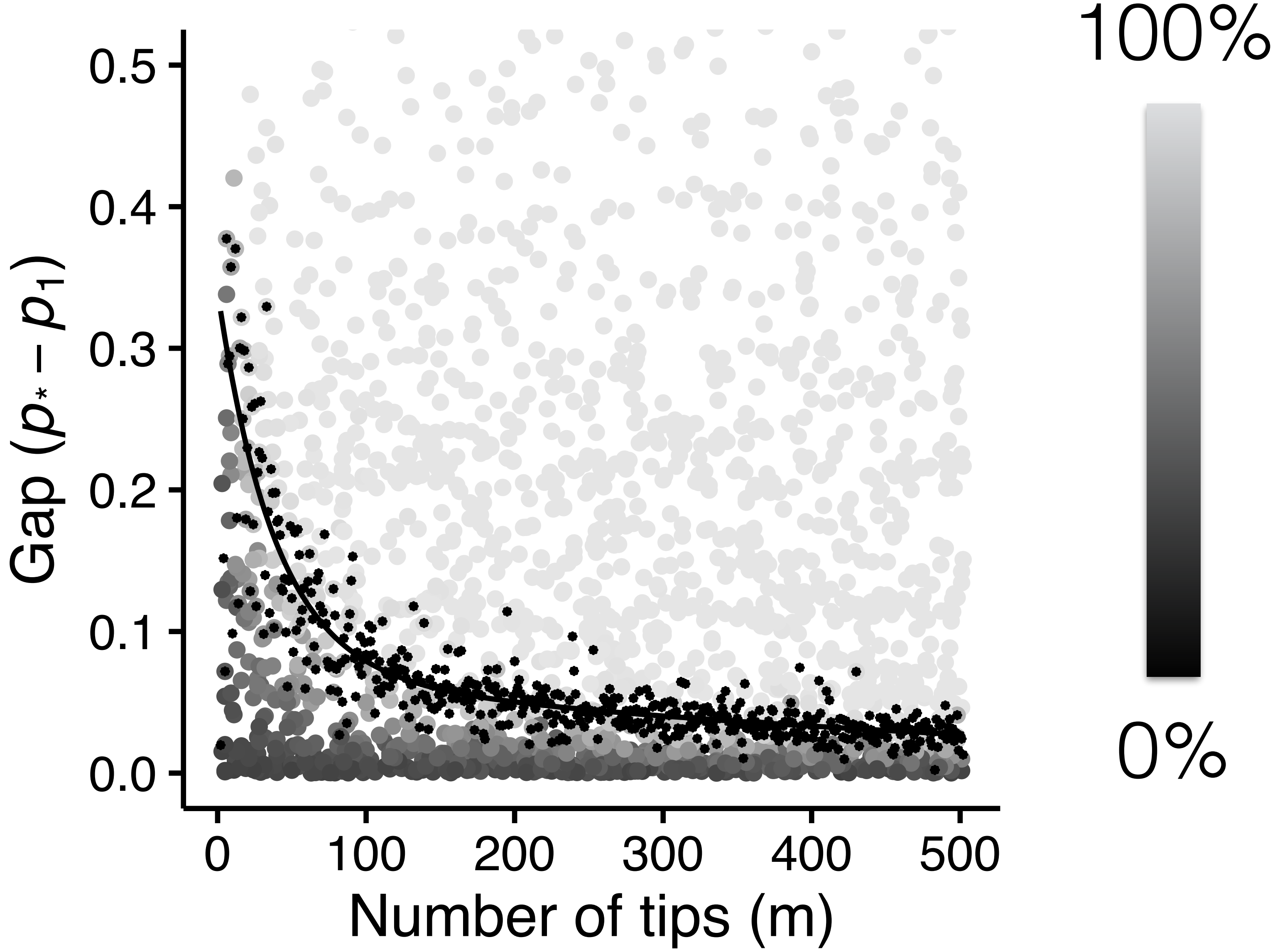
Model Extensions



Geography & Bias

Model Extensions

$$\hat{h} = h_*$$



Experiments

Look promising!

What's next?

Organized Crime & Corruption

Thanks.