

# Wisdom of the Biased Crowd

Despite their biases, crowds predict more accurately than experts

## Introduction

### The Long-Shot Bias



x1.25

VS.



x4

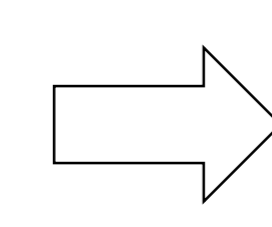
Expected return on bet: **-2%**  $\neq$  Expected return on bet: **-10%**

## Method

### Simple Way to Adjust for the Bias

To improve predictions derived from bookmakers' odds

$$p_i^* = \frac{(p_i)^\alpha}{\sum_j (p_j)^\alpha}$$

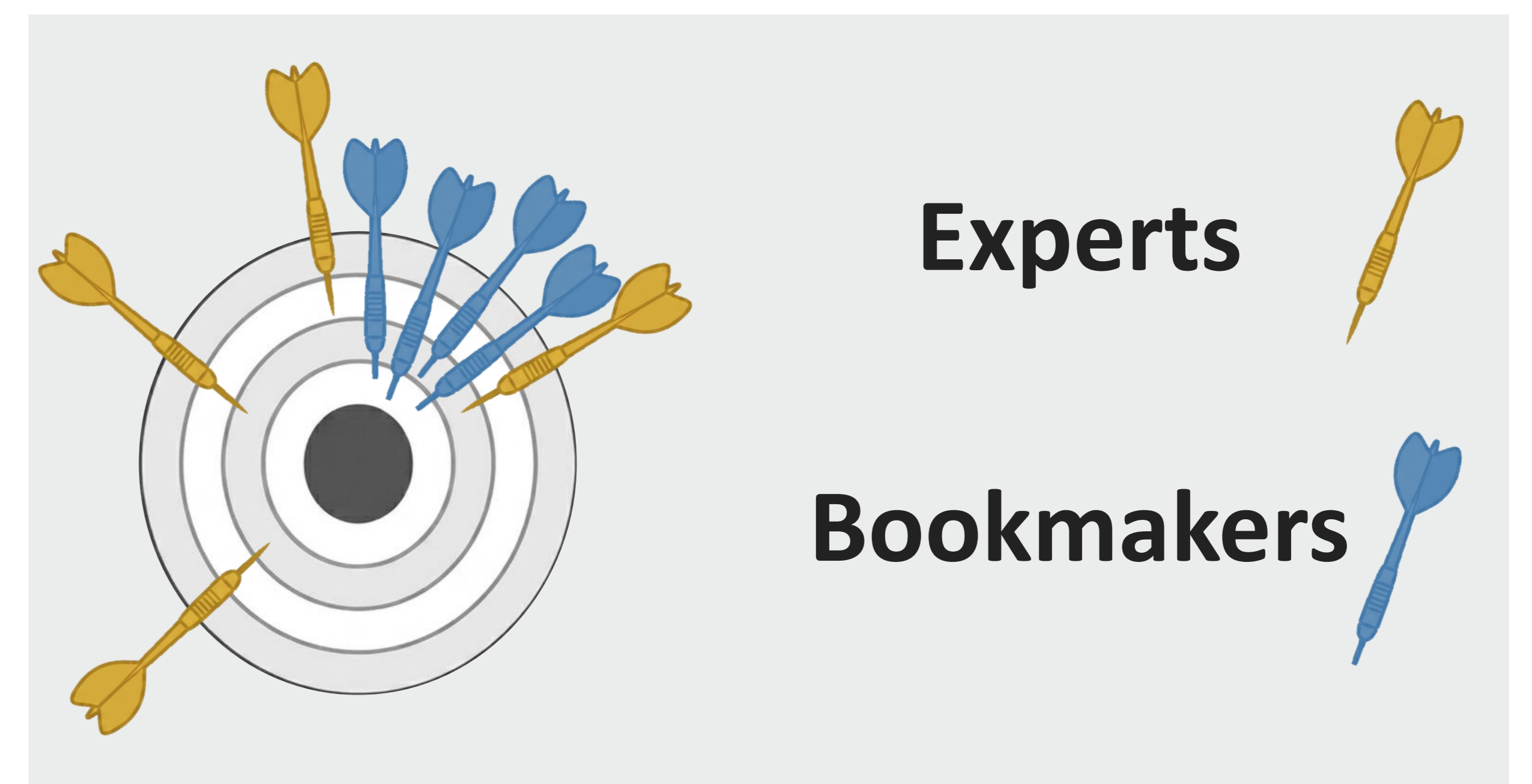
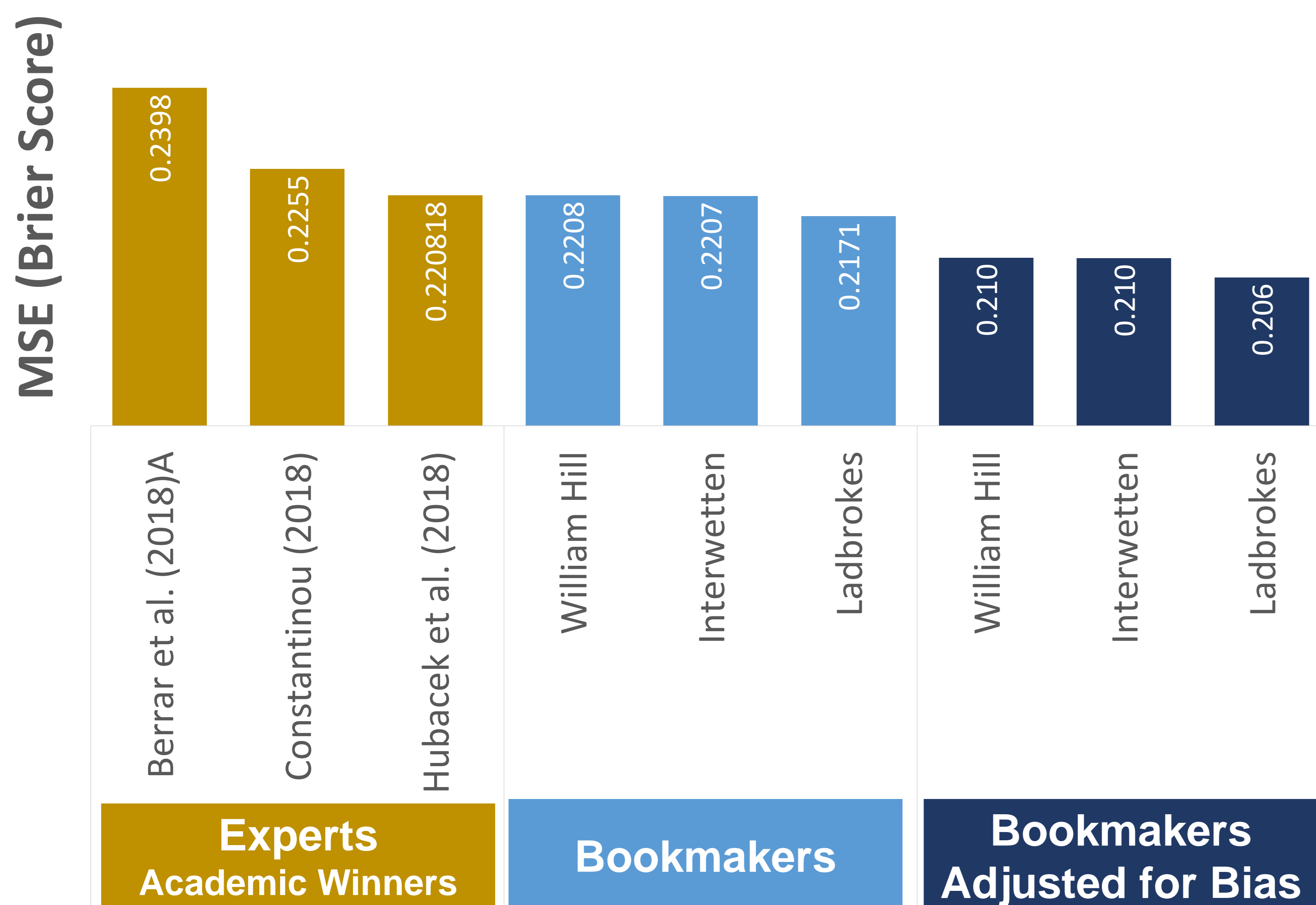


- For  $\alpha > 1$ :
- Decrease low predictions
  - Increase high predictions

Where  $p_i = 1/\text{odds}_i$  is the base prediction.

## Main Result

Bookmakers (The Crowd) predict best – Though can be easily improved



Why are Bookmakers better – even when clearly biased?

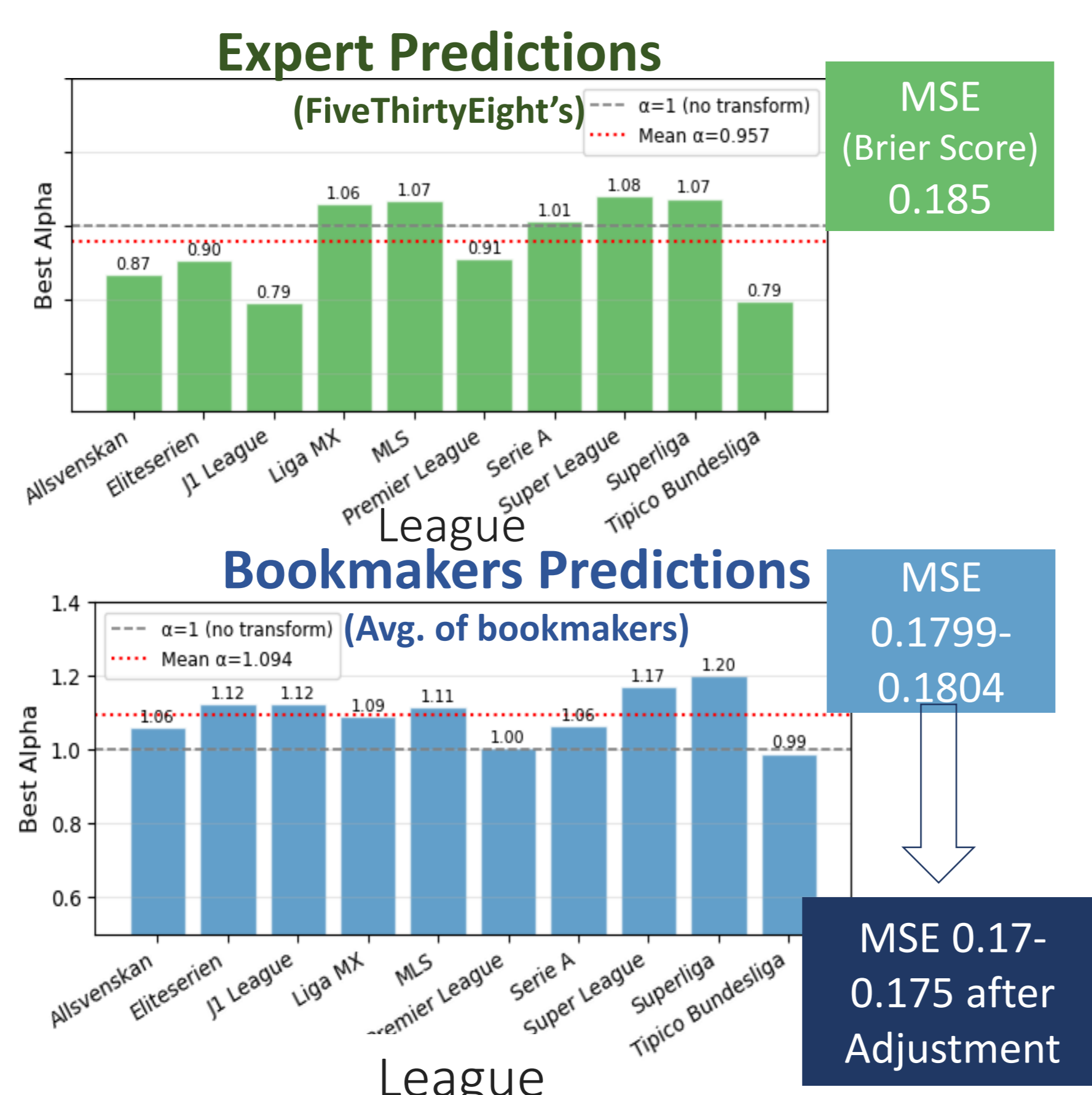
### Two competing explanations

- **A Last-Minute Data Inclusion? – NO**
  - + Closing Odds outperform Opening Odds
  - However, Bookmaker also Win in Calibration Scores
- **Wisdom of the (Biased) Crowd – YES**
  - Predictions MSEs: Exchange < Odds < Experts

Figure. Predictions' MSE Worldwide leagues, 2017

## Results #2 (Some More Experts)

Estimated  $\alpha$ :  
Consistent  
Bias for  
Bookmakers,  
Not Experts



Figures. Worldwide leagues, 2017-2020, 10K matches

### Notes on non-academic / non industry models

Why in recent Kaggle competitions it is more common to predict better than odds?

- Influx of submissions masking pure luck as algorithmic edge
  - o Kaggle Football-match-probability-prediction, 2022 - Bookmaker odds finished 2 out of 381 models
  - o Kaggle march-machine-learning-mania-2025 - 7 of the 8 in the money gambled on multiple games, and more than half of predictions had MSE>0.25 (worse than random baseline).
- Improvement of algorithms. Specifically, XGBoost
  - o Kaggle march-machine-learning-mania – E.g., 2025 winner code wins bookmakers' odds even in 2023 and 2026 data
- Kaggle march-machine-learning-mania-2024 - Incorporation of prediction markets utilities in models.

Does that reduce from the prediction quality of bookmakers?

- During march-machine-learning-mania-2024 submissions were allowed to use odds
  - o From 0.16 MSE to 0.05 MSE – incorporating bookmakers' odds has great predictive value
  - o The 100<sup>th</sup> model there was better than all of 1<sup>st</sup> models in every other year

### Special thanks

To Todd R. Kaplan, and Ori Plonsky for their valuable feedback