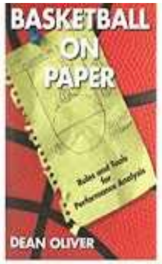
Winning in Basketball with Data and Machine Learning

Konstantinos Pelechrinis
University of Pittsburgh
@kpelechrinis
Math in Sports International Conference
Athens, Greece, 2019

A brief history of…basketball analytics

WINVAL

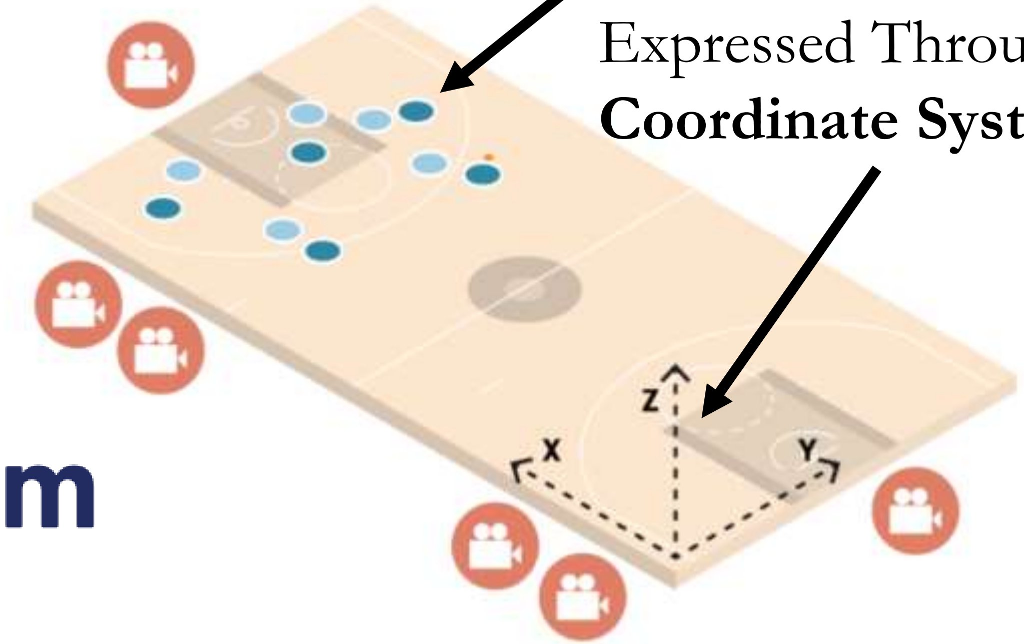


1989 1995 2002 2004 2006 2012

Tracking data

STATS

Second Spectrum

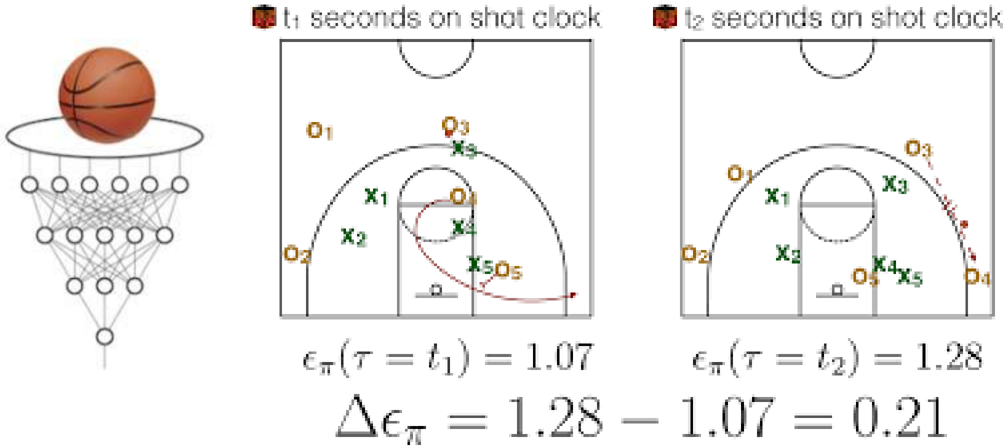


Court Locations
Expressed Through
Coordinate System

Tracking data

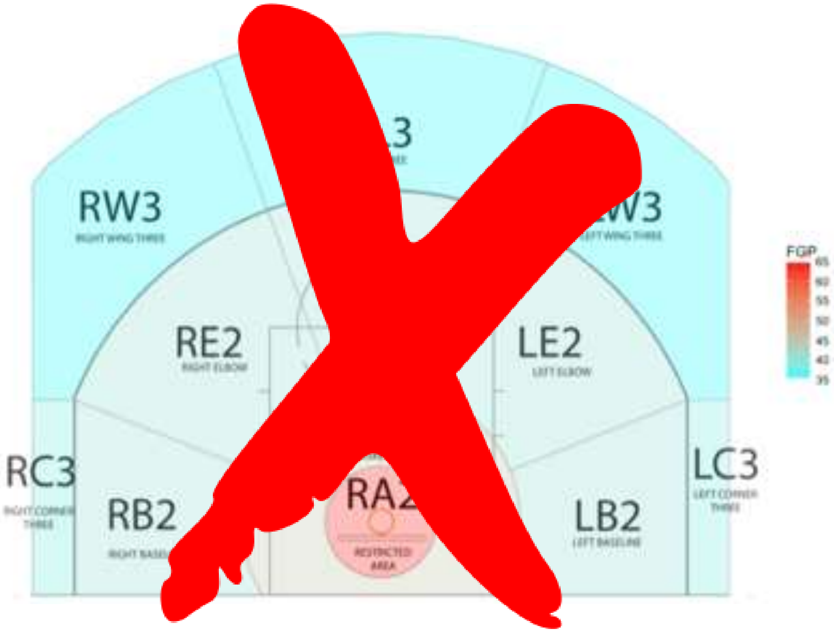
- Not a panacea but can help fill gaps in the current analytics
  - Allows to measure aspects of the game that are closely aligned with what coaching staff talk and think about
    - E.g., spacing
  - Defensive metrics
  - Shooting ability - beyond FG% (shot quality)
  - …

Roadmap

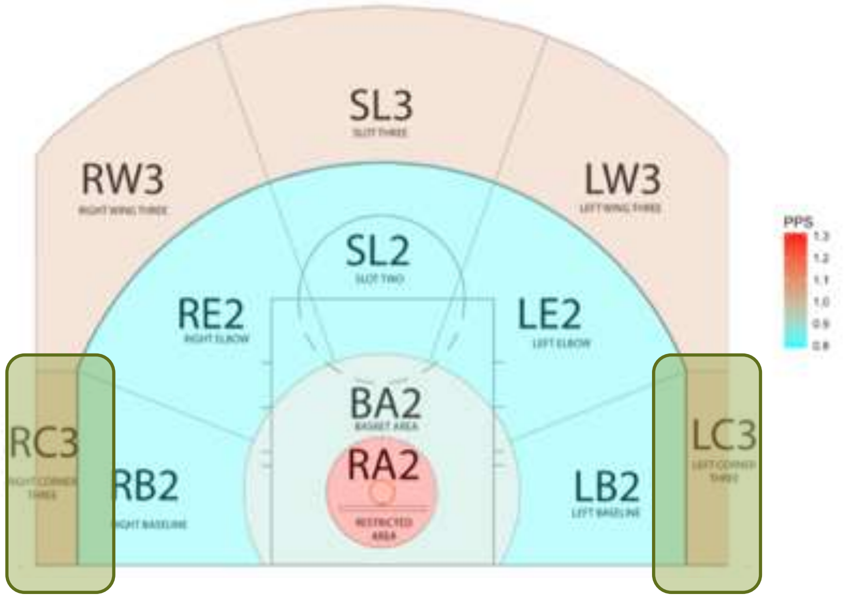


t_1 seconds on shot clock
t_2 seconds on shot clock
$\epsilon_\pi(\tau = t_1) = 1.07$
$\epsilon_\pi(\tau = t_2) = 1.28$
$\Delta\epsilon_\pi = 1.28 - 1.07 = 0.21$

The two-faced *shooting* landscape



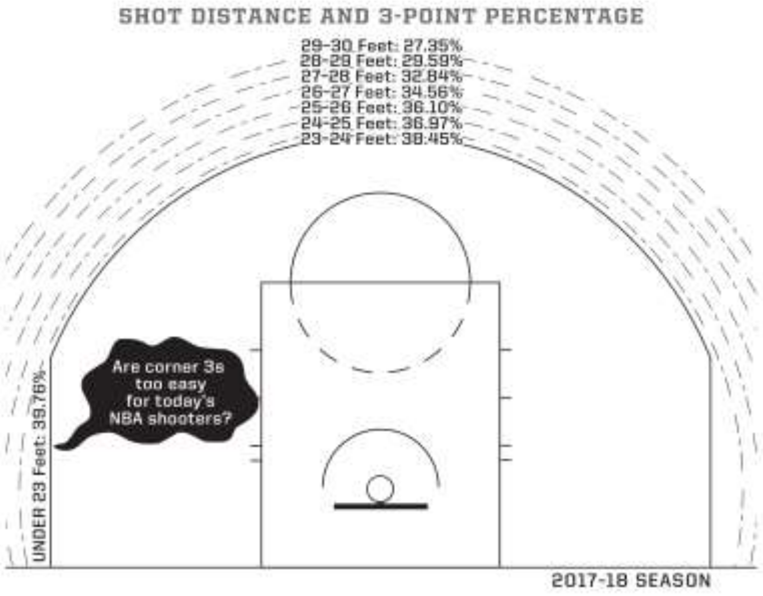
The *make* landscape

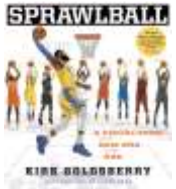


The *efficiency* landscape

More than a decade ago, San Antonio Spurs head coach Gregg Popovich discovered the most valuable 21 inches on an NBA basketball court, and nothing has been the same since.

Those 21 little inches, the difference between a 22-foot corner three and the 23'9" distance at every other spot on the arc, changed everything about the Spurs' offensive attack. And as has long been the case with NBA innovation, the rest of the league eventually caught on to what Popovich and San Antonio were doing.

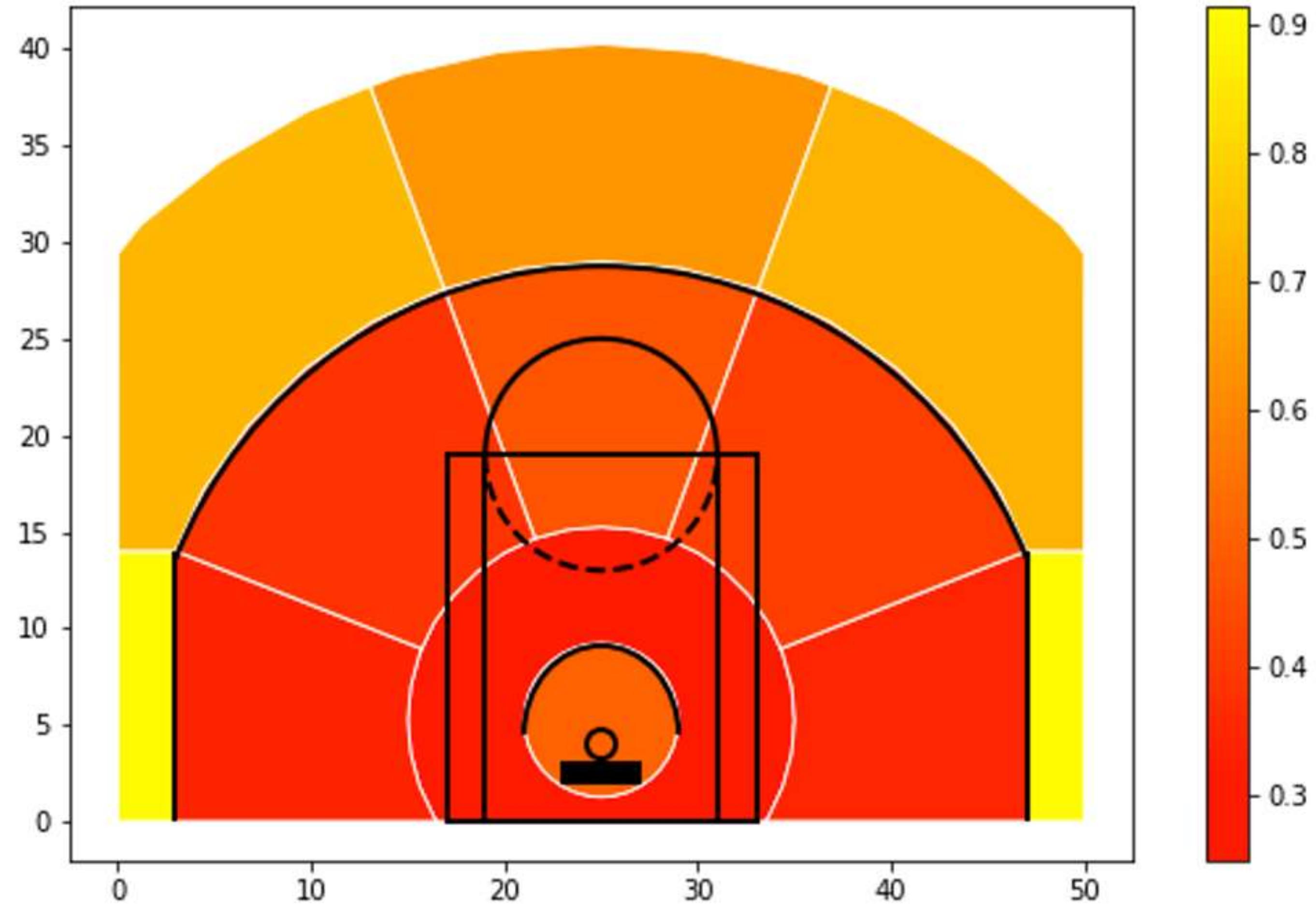


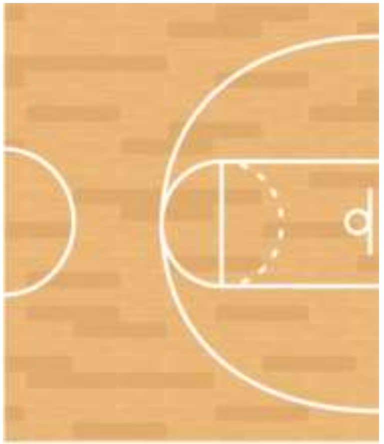


https://www.espn.com/nba/story/_/id/26633540/the-nba-obsessed-3s-let-fix-thing

	Actual Field Goal Percentage (from raw data)	Expected Field Goal Percentage (from logistic model)	 (feet)
	38.7%	35.4%	6.4
	34.7%	33.8%	5.9

Corner 3s are assisted at a much higher rate compared to any other shot

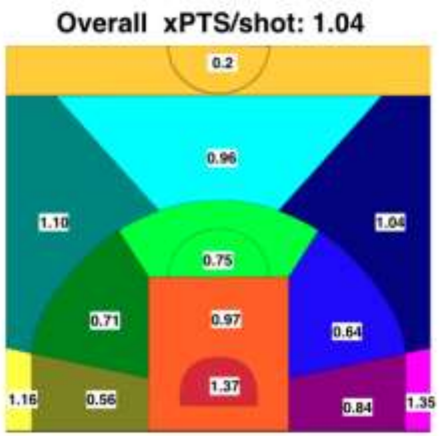




Short corners

Equidistant

Natural experiment



Corner 3s are assisted twice as frequently as the wing 3s & 33% more frequently compared to the top of the key

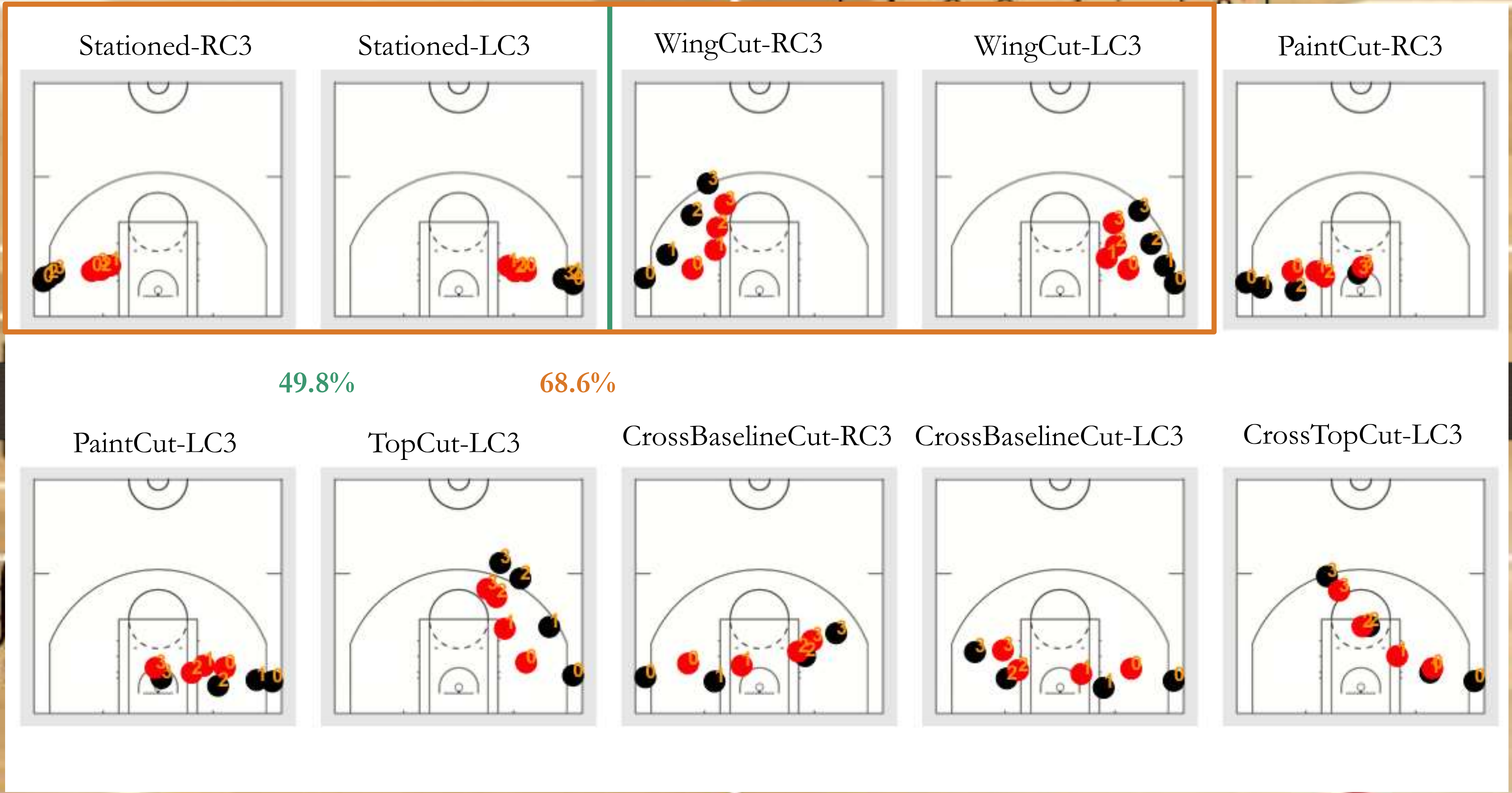
https://theconversation.com/data-reveals-the-value-of-an-assist-in-basketball-113893

Objectives

- Understand how corner 3s are created
  - Focus on the last few moments prior to the shot
  - Are there distinct patterns?
- Focus on assisted shots
- How should the defense react?

Shooter-defender *choreography*

- 2,839 total corner 3s in our dataset from the 2016-17 regular season
  - 90% of them assisted (or potentially assisted) – i.e., 2,559 shots
- Cluster the whole trajectory
  - Feature vector is the whole trajectory of the shooter and the defender for the last 4 seconds of the offense
  - Gap heuristic for number of clusters


Stationed-RC3
Stationed-LC3
WingCut-RC3
WingCut-LC3
PaintCut-RC3
49.8%
68.6%
PaintCut-LC3
TopCut-LC3
CrossBaselineCut-RC3
CrossBaselineCut-LC3
CrossTopCut-LC3

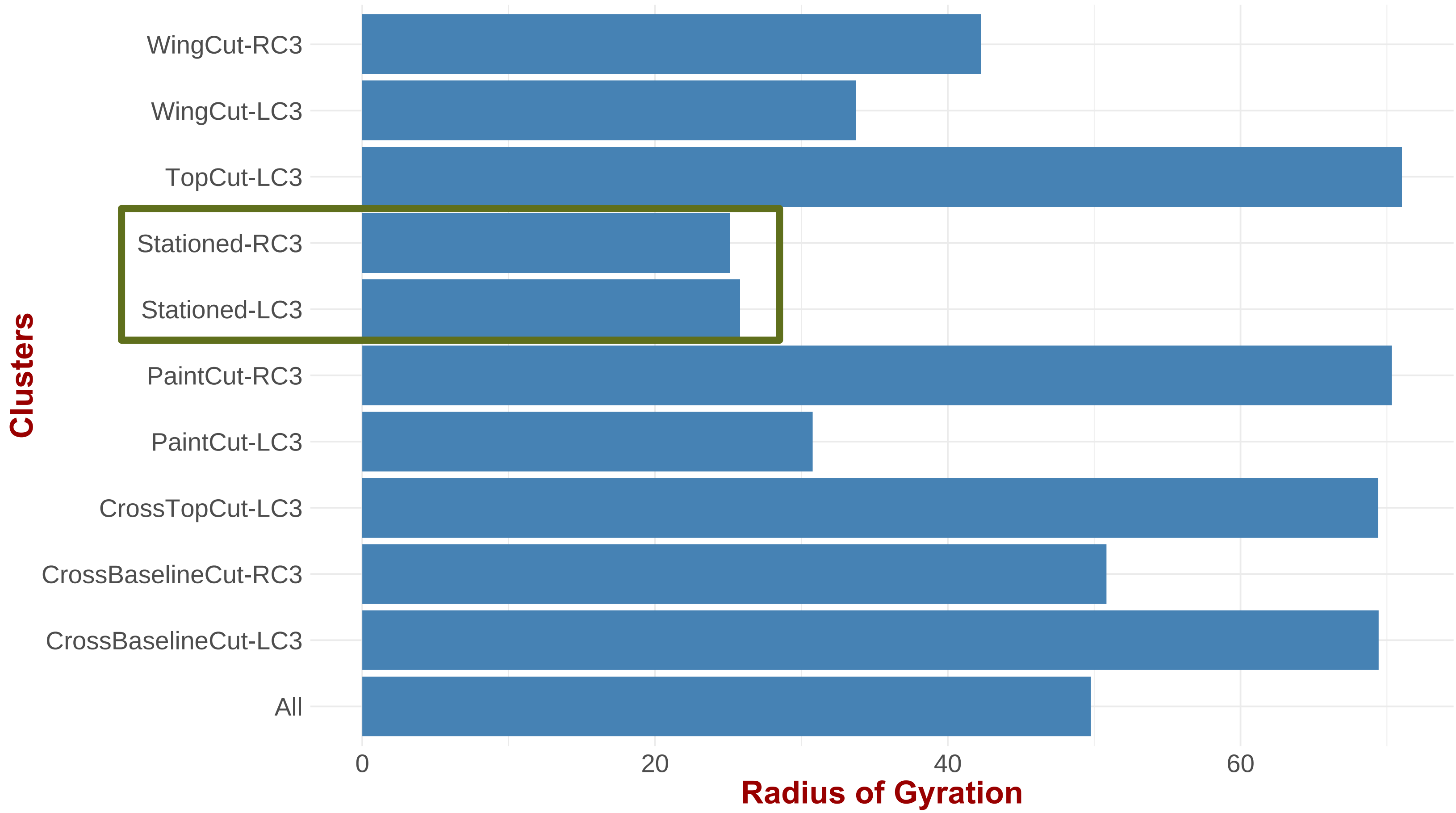
Cluster variance

- Radius of gyration
  - Quantifies the *coherence* of a cluster:

$$r = \sqrt{\frac{1}{N}\sum_{i=1}^{N} dist(\boldsymbol{x}_i, \boldsymbol{x}_{cent})^2}$$

- Radius of gyration measures the standard deviations of distances between the cluster members possessions and the clusters centroid
  - Low radius of gyrations → members of a cluster are *near* the centroid

- Stationed corner 3s are also the least variable!



What did we learn?

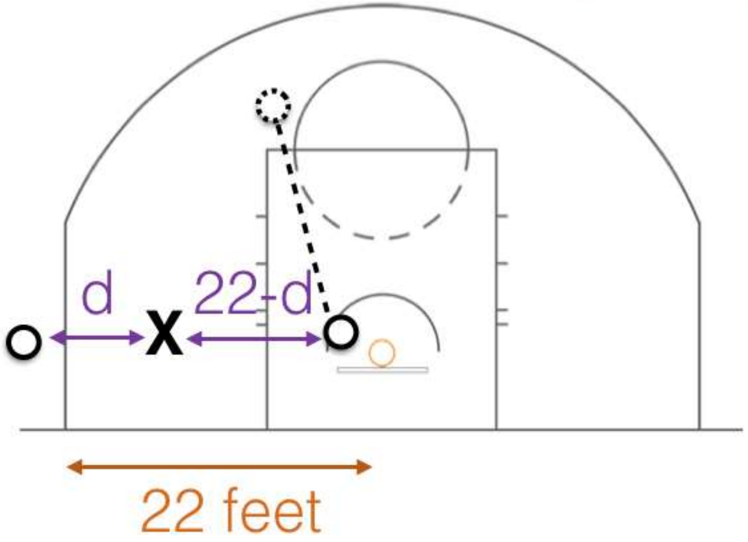
- Almost half of the corner threes involve the shooter stationed at the corner and the defender being indecisive between covering him and protecting the paint
- It seems that the dilemma from a defensive strategy perspective is: move towards the ball or stay close to the corner?
  - Can we make game theoretic arguments to obtain an answer?

Zero-sum game

Defense strategy:
distance to the corner

$d \in \{1,2,3,\ldots,21\}$

Offense strategy:
{drive, pass-corner3}



$$
\begin{array}{cc}
 & \textbf{Offense} \\
\textbf{Defense} & \begin{bmatrix} \pi_{1,1} & \pi_{1,2} \\ \cdots & \cdots \\ \pi_{21,1} & \pi_{21,2} \end{bmatrix}
\end{array}
$$

Payoff matrix

When offense chooses kick-out pass

Quantifying Shot Quality in the NBA

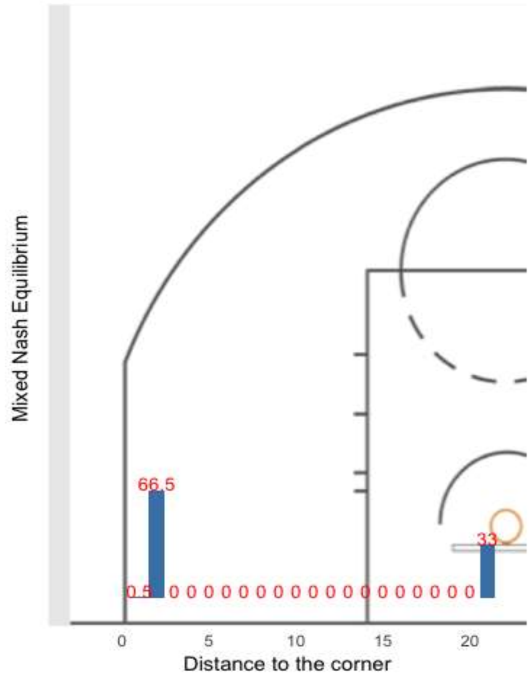
Yu-Han Chang, Rajiv Maheswaran, Jeff Su, Sheldon Kwok, Tal Levy, Adam Wexler, Kevin Squire
Second Spectrum, Inc.
Los Angeles, CA
ychang@secondspectrum.com

When offense chooses drive

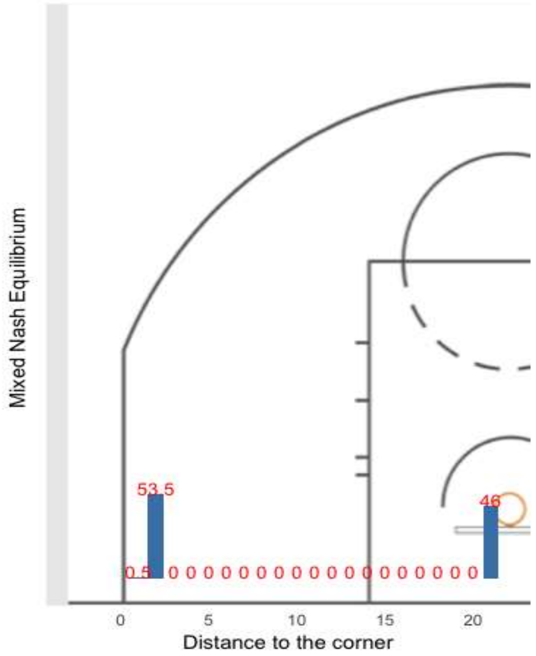
$$\boxed{\pi_b} \cdot \left(1 - \frac{1}{\boxed{a}^{(22-d)}}\right)$$

Impact of double team

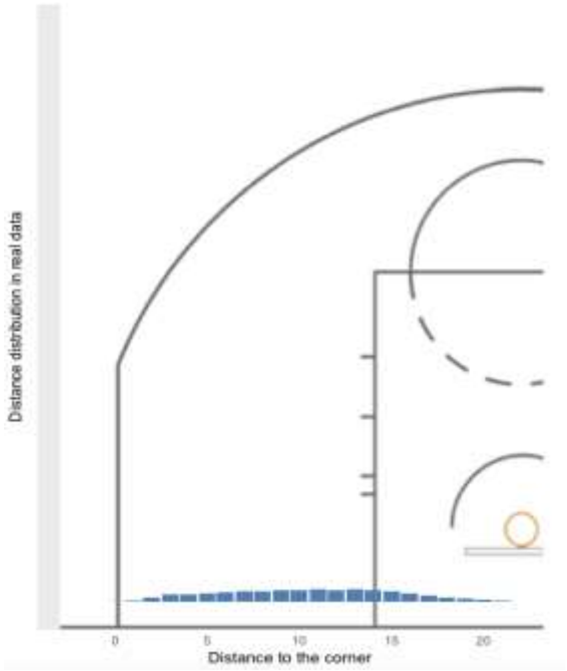
Impact of primary defender



α = 1.3



α = 1.9



Actual distribution

Obtained only from corner 3 instances

Analytics-driven Sixers ride the numbers to NBA playoffs

Posted: April 18, 2018 - 11:22 AM

Marcus Hayes | @inkstainedretch | mhayes@inquirer.com

CHARLES FOX / STAFF PHOTOGRAPHER

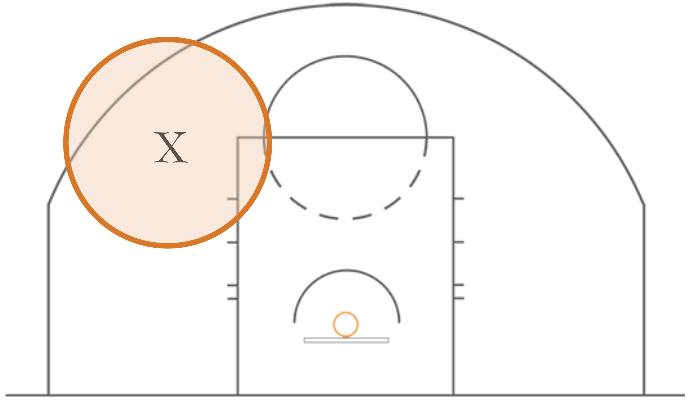
Trust the Process, if you like.

The Sixers trust the numbers.

"To be fair, some of the information and actions are counterintuitive. For instance: when you play a team with deadly corner three-point shooters, it's smarter to stay on your man in the corner and not help when an opponent drives down the lane, and that can look foolish for the unwashed."

Connections with behavioral science

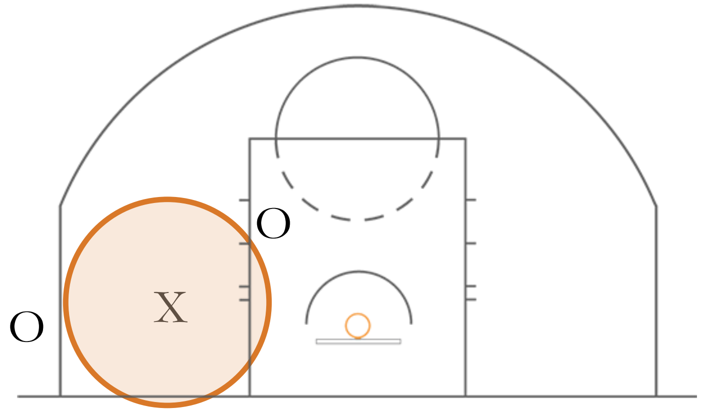
- Sports have been a testbed for behavioral economics and decision making sciences
  - The defender's choice can provide insights on how humans assess spatial risk



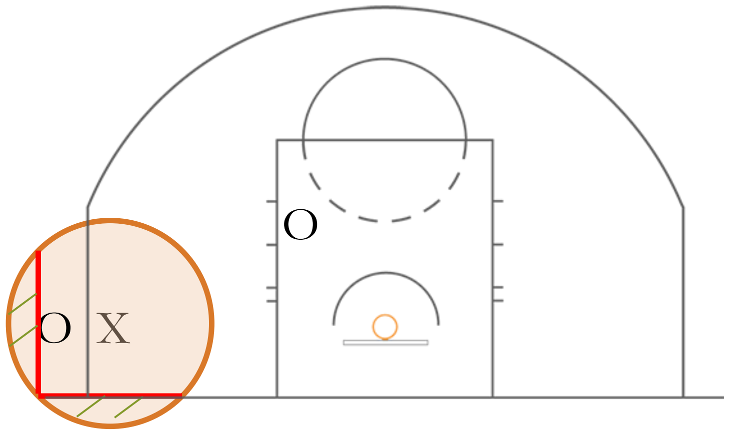
Area of defensive influence

Unit disk model for the area of defensive influence: The defensive player has defensive influence in an area of radius r around his location

X: Defensive player
O: Offensive player



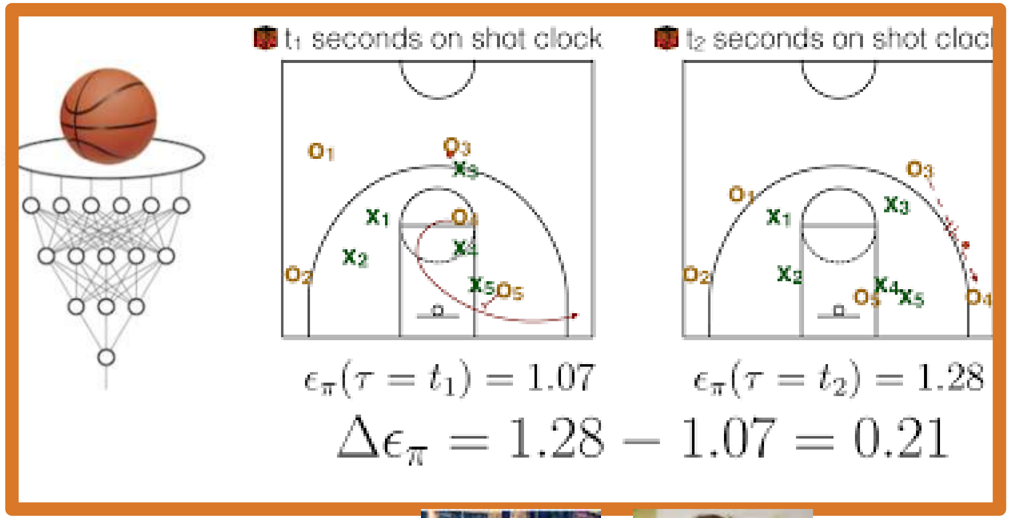
Lingering between the driving lane and the corner 3 allows the defender maximize the overlap between his defensive influence area and the court



If the defender commits at the corner, his area of influence extends outside the court! He assumes his defensive influence is smaller.

But not every area on the court is equally *important* and defenders seems to *ignore* that.

Roadmap


t_1 seconds on shot clock
t_2 seconds on shot clock
$\epsilon_\pi(\tau = t_1) = 1.07$
$\epsilon_\pi(\tau = t_2) = 1.28$
$\Delta\epsilon_\pi = 1.28 - 1.07 = 0.21$

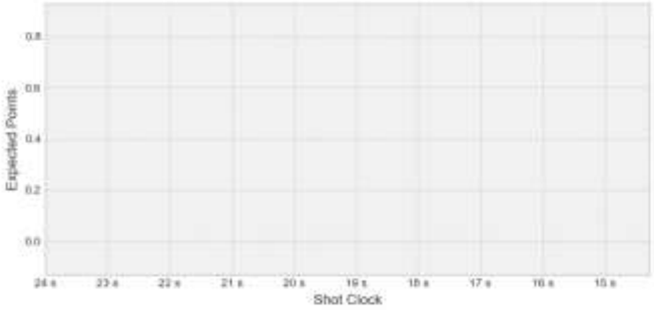
Evaluating Micro-Actions? What and Why?

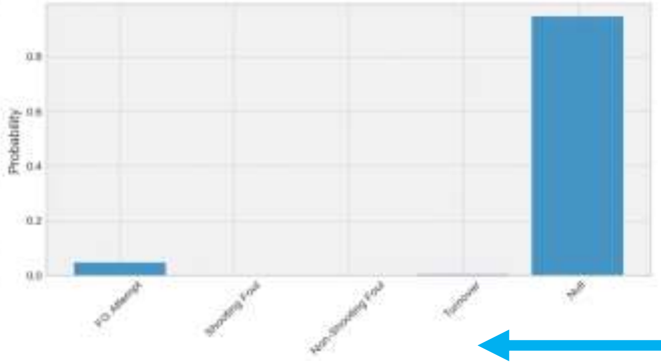
- **Micro-Actions** are individual moves taken by players
  - Pass, Screen, etc.
- Combine to form overall strategy
- What is a particular action's value?
- **Goal:** Quantify value of actions in terms of *expected points added*
  - **Why?** Provides players and coaches an **interpretable** tool to improve

Example: How much value does O5's screen add for the offense?

Evaluating Micro-Actions? How?

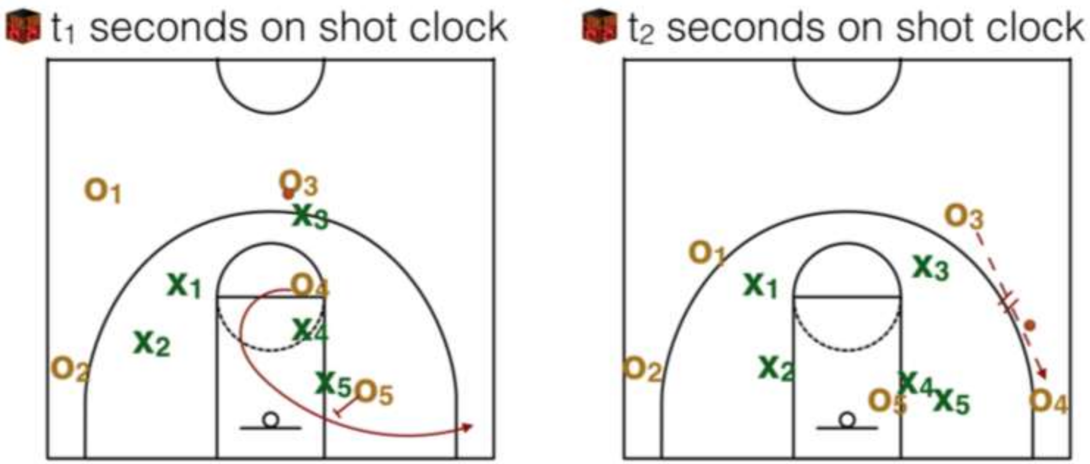
To identify value we predict the **expected points** to be scored by the offense during a possession at *each moment*





Based on **probability** of different *possession outcomes* at each moment

Evaluating Micro-Actions? How?



$\epsilon_\pi(\tau = t_1) = 1.07$ $\quad$ $\epsilon_\pi(\tau = t_2) = 1.28$

$$\Delta\epsilon_\pi = 1.28 - 1.07 = 0.21$$

Predicting Expected Points

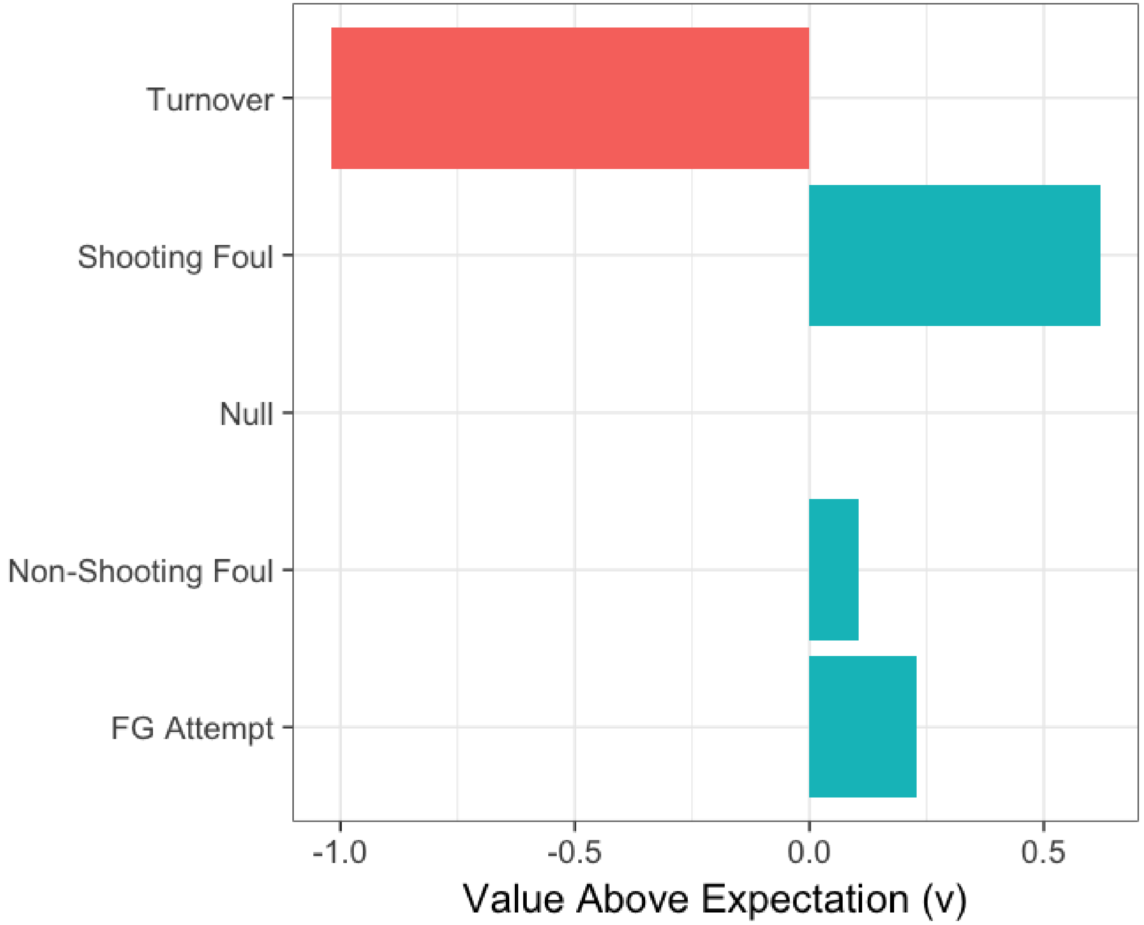
(3) Assume average point-value

(4) Modify based on probabilities of outcomes

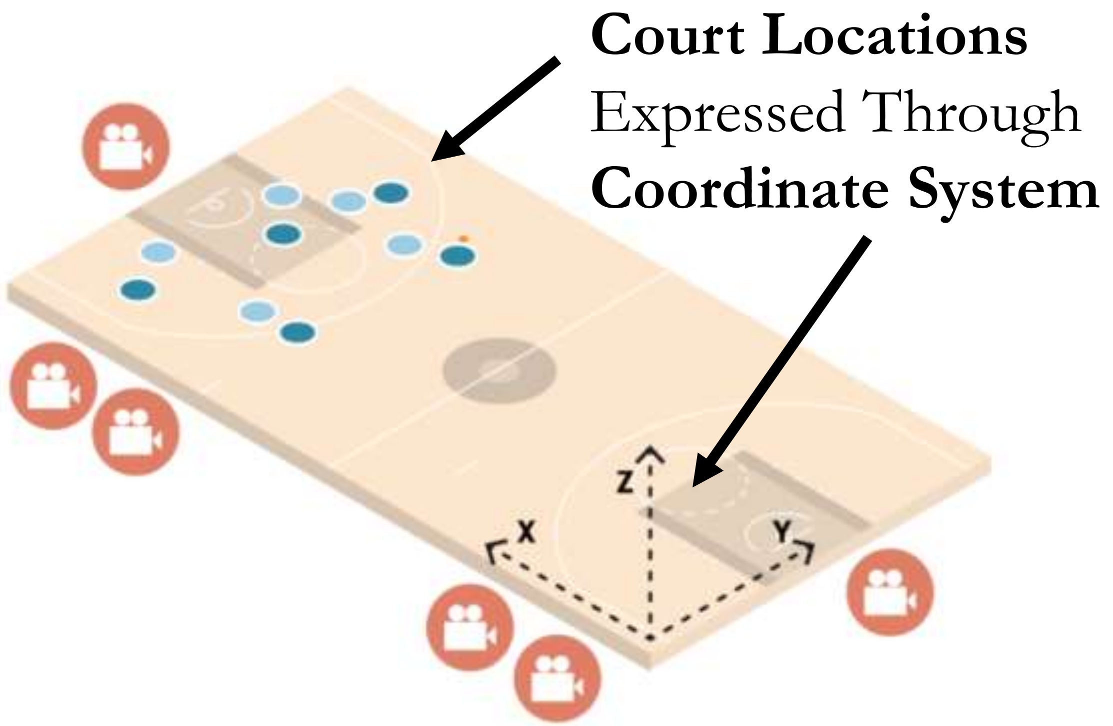
$$\epsilon_{\pi}{}^{(i)}(\tau) = \beta_{\pi} + \mathbb{E}_{y \sim \boxed{P(y)}}[\nu(y) | \underbrace{W_{\tau}^{(i)}, \{s_j^{(i)}\}_{j=1}^{10}}]$$

(2) Learn the probability of outcomes at each moment

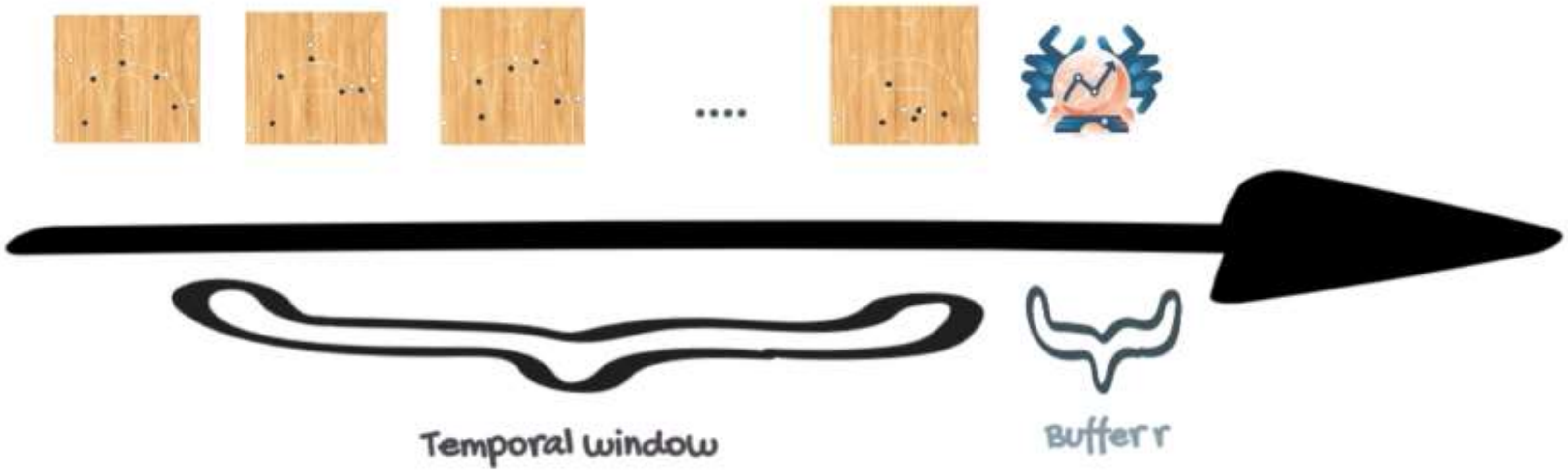
(1) Using data from NBA games


Turnover
Shooting Foul
Null
Non-Shooting Foul
FG Attempt
-1.0
-0.5
0.0
0.5
Value Above Expectation (v)

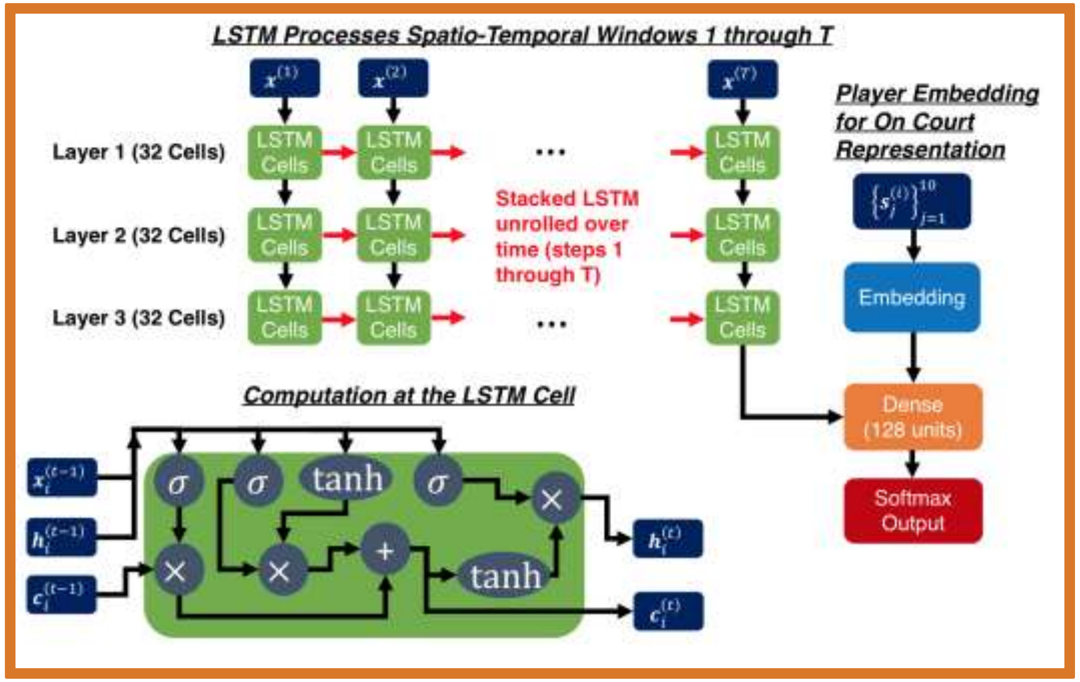
- 750 games from NBA (2016-17)
  - Over 134,000 total possessions
- Player and ball **court locations** provided **25 times** per second
- Highly annotated!



DeepHoops at a glance



DeepHoops architecture



Network predicts outcome probabilities based on **weights** (to be learned)

$$i_i^{(t)} = \sigma(b_i' + \sum_j U_{i,j}' x_j^{(t)} + \sum_j W_{i,j}' h_j^{(t-1)})$$

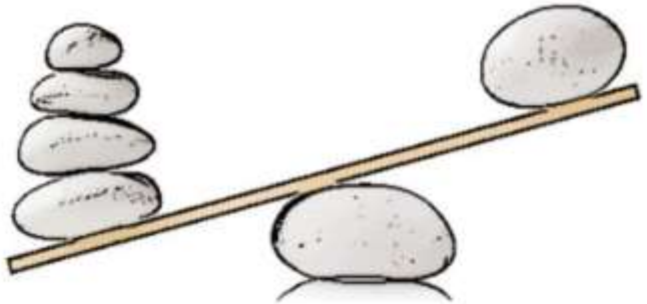
$$f_i^{(t)} = \sigma(b_i^f + \sum_j U_{i,j}^f x_j^{(t)} + \sum_j W_{i,j}^f h_j^{(t-1)})$$

$$c_i^{(t)} = f_i^{(t)} c_i^{(t-1)} + i_i^{(t)} \tanh(b_i^c + \sum_j U_{i,j}^c x_j^{(t)} + \sum_j W_{i,j}^c h_j^{(t-1)})$$

Severe imbalance

Single possession

1. Single non-null sequence
2. Several null sequences

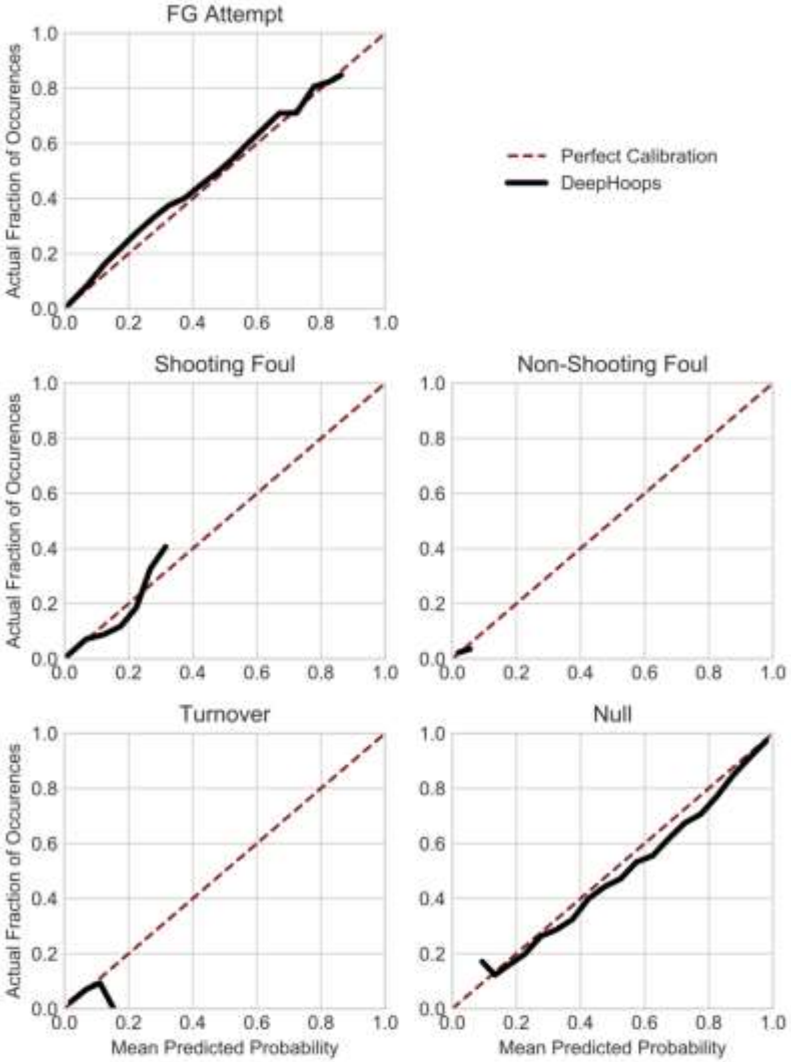


Down sampling the `null` class

For every possession

- Keep the final sequence
- Uniformly sample K (null) time points

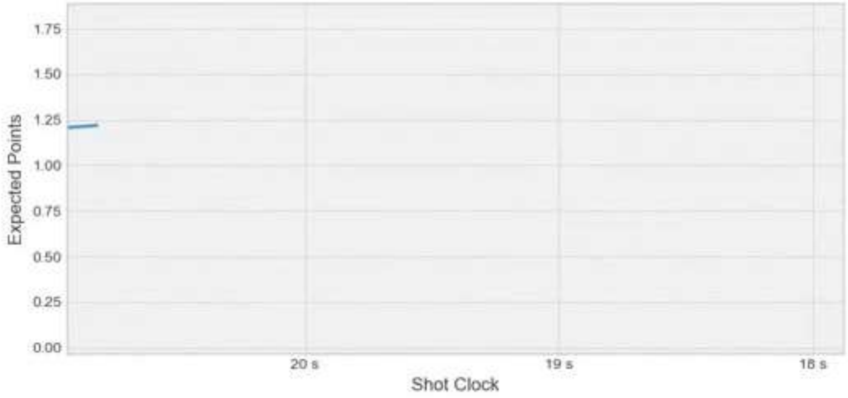

Experiments

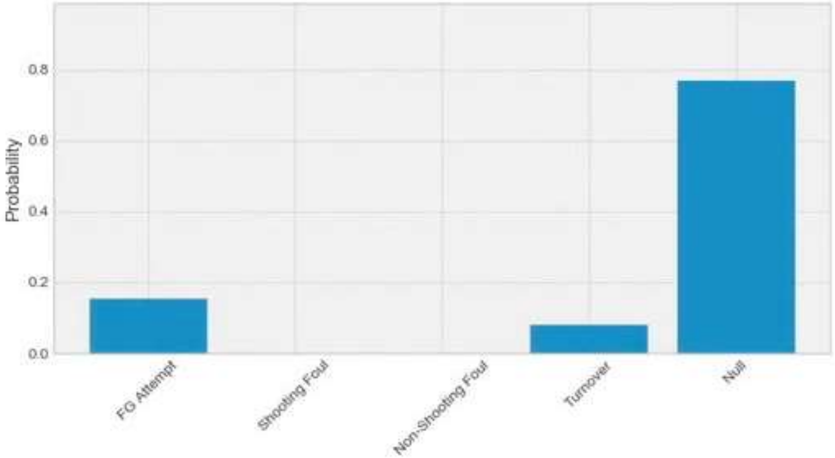


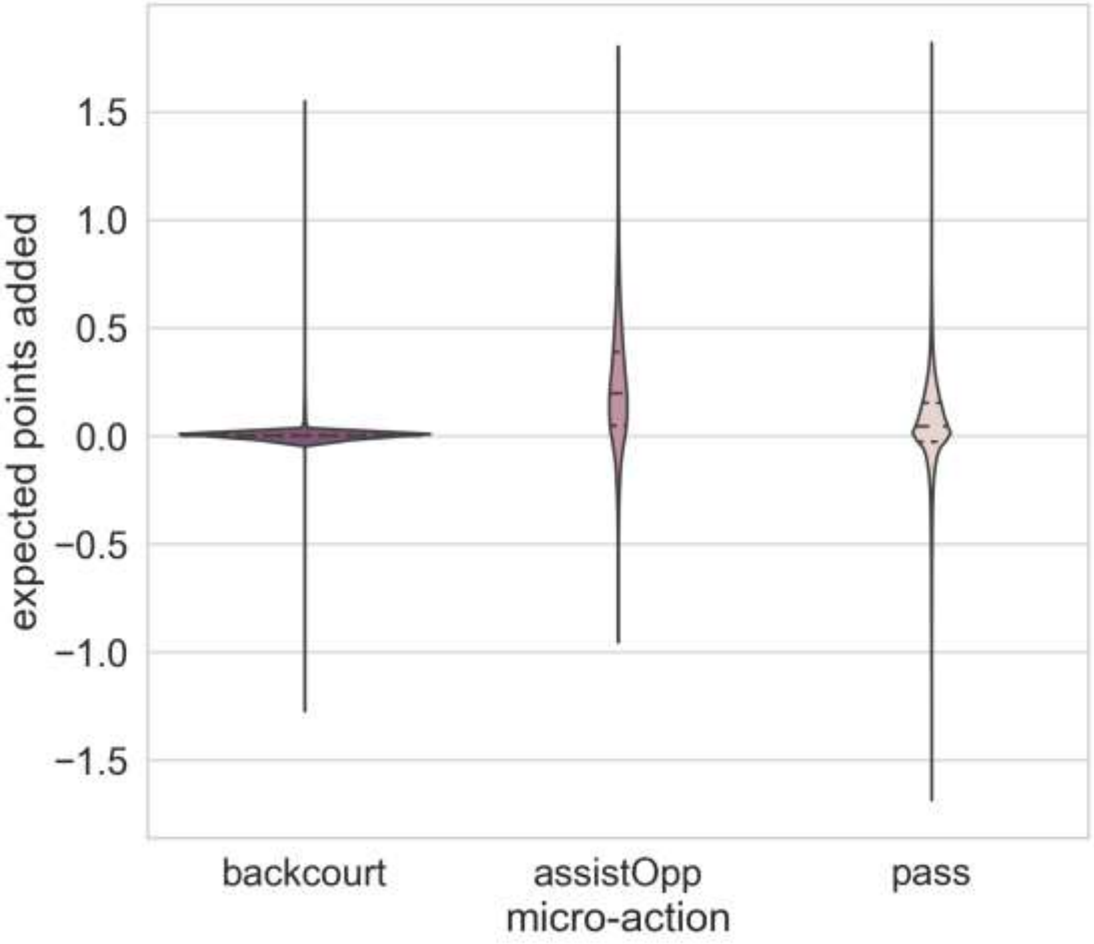
FG Attempt
Shooting Foul
Non-Shooting Foul
Turnover
Null
Actual Fraction of Occurences
Mean Predicted Probability
Perfect Calibration
DeepHoops

Expected Points Added by a Pass

Did the expected points increase at the end of the pass?






expected points added
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
backcourt
assistOpp
pass
micro-action

Parting thoughts

- Analytics are here to stay in basketball
- New tech developments will make it easier for more leagues to utilize similar tech
- Always a new edge to gain

http://www.github.com/anthonysicilia/DeepHoopsRealTimeApplication