Statistics in Sports: Basketball Overview

ZACHARY BINNEY, PHD MPH

OXFORD COLLEGE OF EMORY UNIVERSITY

FALL 2021

Catch-up

- Soccer example of the shift (Daolong)
- Zach's article on SIERA

Roadmap

- 1. Sports Analytics “Culture Wars”
- 2. Rate Stats
- 3. Expected (Shot) Values
  - High-Value and Low-Value Shots (a/k/a Death of the Midrange Jumper)
- 4. The Quest for Holistic Statistics
  - Four Factors
  - PER
  - Plus-Minus
- 5. The Future: Expected Points Value Added (EPVA)
- 6. Hot Hand Fallacy (?) and Selection Effects

Sports Analytics “Culture Wars”

Analytics Culture War

- "Outsiders" not just in a nerd vs. jock sense
  - In NBA and NFL (in U.S.) especially, also a racial/diversity component
  - Think about how you come across in the environment you're entering
  - Watchword: humble
- Jalen Rose, Bomani Jones, Shane Battier, and basketball analytics "resistance":

Jalen Rose, on NBA Analytics:

No. 1, there are many people that feel like it has a cultural overtone to it that basically suggests that, even though I may not have played and you did, I am smarter than you, and I know some things that you don't know, and the numbers support me, not you. Two, you notice that, when it is a powerful job in sports—whether it is an owner, whether it is a president, whether it is a general manager, whether it is a coach—usually in football and basketball, sports that are primarily dominated by black Americans, it's also an opportunity to funnel jobs to people by saying that, "I am smarter than you because the numbers back up what I say, and I am more read. I study more. I am able to take these numbers and manipulate my point." It's almost like when you hear that a player doesn't have experience at doing X job. People that normally get the jobs you are describing don't, either. They didn't play at most levels, but that suffices as their "experience" and validates their opportunity for power.

Just to be clear, when you say "cultural overtones," you mean racial overtones?

Correct. And one other point I want to make with that: it is laughable to me when

Source: SEE links

Analytics Culture War

Source: https://twitter.com/bomani_jones/status/1137124576100569095?lang=en

Rate Stats

Which Player Was Better?

- Stats from a 5-game playoff series:
- Player A:

Player	Age	G	GS	[redacted]	FG	[redacted]	3P	[redacted]	FT	[redacted]	PTS	[redacted]
[redacted]	27	5	5	[redacted]	43	[redacted]	7	[redacted]	46	[redacted]	139	[redacted]

- Player B

Player	Age	G	GS	[redacted]	FG	[redacted]	3P	[redacted]	FT	[redacted]	PTS	[redacted]
[redacted]	27	5	5	[redacted]	52	[redacted]	4	[redacted]	31	[redacted]	139	[redacted]

Source: basketball-reference.com

Which Player Was Better?

- Stats from a 5-game playoff series:
- Player A:

Player	Age	G	GS	MP	FG	[redacted]	3P	[redacted]	FT	[redacted]	PTS	[redacted]
[redacted]	27	5	5	189	43	[redacted]	7	[redacted]	46	[redacted]	139	[redacted]

- Player B

Player	Age	G	GS	MP	FG	[redacted]	3P	[redacted]	FT	[redacted]	PTS	[redacted]
[redacted]	27	5	5	204	52	[redacted]	4	[redacted]	31	[redacted]	139	[redacted]

Source: basketball-reference.com

Which Player Was Better?

- Stats from a 5-game playoff series:
- Player A:

Player	Age	G	GS	MP	FG	FGA	3P	3PA	FT	FTA	[redacted]	PTS	[redacted]
[redacted]	27	5	5	189	43	90	7	20	46	57	[redacted]	139	[redacted]

- Player B

Player	Age	G	GS	MP	FG	FGA	3P	3PA	FT	FTA	[redacted]	PTS	[redacted]
[redacted]	27	5	5	204	52	124	4	18	31	41	[redacted]	139	[redacted]

- Who are Player A and Player B?
  - Bonus points: what series?

Source: basketball-reference.com

Which Player Was Better?

- Stats from 2012-13 MIA-NYK 1st round matchup:
- Player A:

Player	Age	G	GS	MP	FG	FGA	3P	3PA	FT	FTA	[redacted]	PTS	[redacted]
LeBron James	27	5	5	189	43	90	7	20	46	57		139	

- Player B

Player	Age	G	GS	MP	FG	FGA	3P	3PA	FT	FTA	[redacted]	PTS	[redacted]
Carmelo Anthony	27	5	5	204	52	124	4	18	31	41		139	

Source: basketball-reference.com

Denominators in Basketball

- From **counting** to **rate** stats in basketball:
  - Shooting: **Field Goal Percentage (FG%)** = $\frac{Field\ Goals\ Made\ (FGM)}{Field\ Goal\ Attempts\ (FGA)}$
  - What about other stats? Let's take *team points* as an example…

Denominators in Basketball

- Points per…
  - Game?
  - Minute (or “Per 36” or “Per 48”)?

Denominators in Basketball: Possessions

- **Possessions:**

> The basketball possession is the equivalent of an atom in science. It's a basic building block from which an entire field of analysis was born.
>
> A possession is named as such because it marks the entire time a team possesses the ball. This should not be confused with plays, which are separated by shot (or free throw) attempts and turnovers. Offensive rebounds extend possessions, not plays. Note: not everyone calls that a "play" but it's an important distinction to make with offensive rebounds.

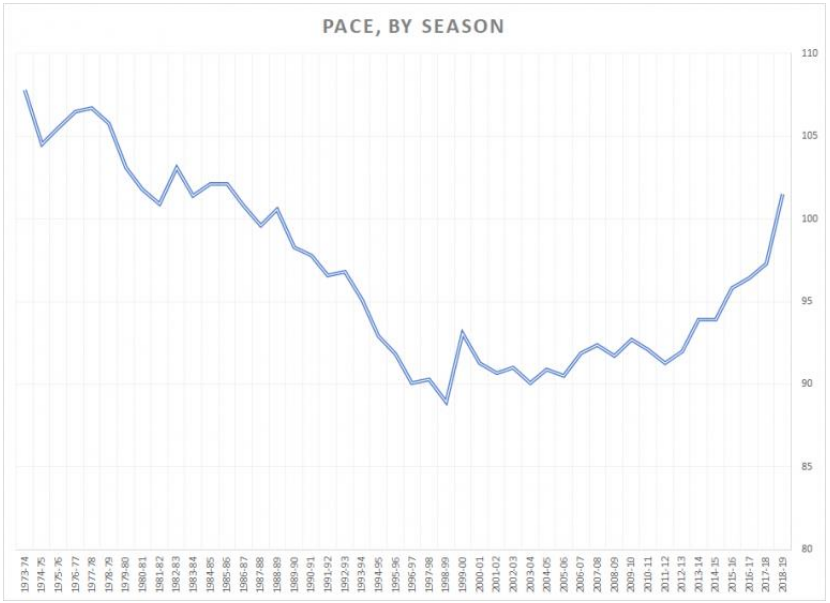
- Estimated via formula. Several versions, but simple one is

$$\text{Possessions} = [FGA - Offensive\ Rebounds\ (OREB)] + Turnovers\ (TOV) + 0.44 * Free\ Throw\ Attempts\ [FTA]$$

Source: https://fansided.com/2015/12/21/nylon-calculus-101-possessions/

Denominators in Basketball: Possessions

- Why possessions?
  - 1. Is it easier to score 95 points in a 100-possession game or a 120-possession game?
  - 2.



  - Many basketball stats are standardized to **per 100 possessions**

Source: https://www.lineups.com/articles/why-nba-game-pace-is-at-historic-high/

(Team) Offensive and Defensive Rating

- **(Team) Offensive Rating (ORtg) or Efficiency (OFF EFF)** = $\frac{Points\ Scored\ (Pts)}{Possessions} x100$
- **(Team) Defensive Rating (DRtg) or Efficiency (DEF EFF)** = $\frac{Points\ Allowed\ (PA)}{Possessions} x100$

2020-2021 Hollinger Team Statistics

Season: 2020-2021 Regular Season

Team Projections

Player Statistics »

Hollinger Stats - Offensive Efficiency

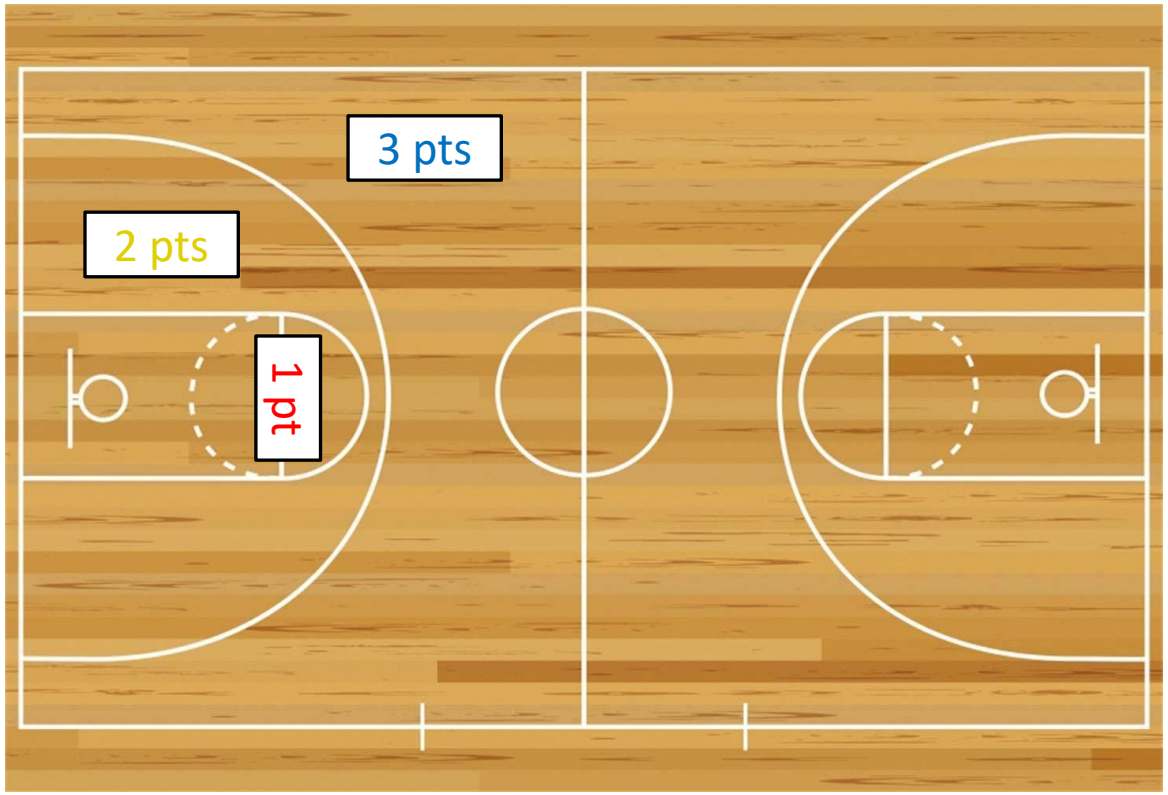
RK	TEAM	PACE	AST	TO	ORR	DRR	REBR	EFF FG%	TS%	OFF EFF	DEF EFF
1	Brooklyn	101.8	19.5	12.2	21.4	77.6	50.8	57.5	61.0	**116.3**	110.6
2	LA Clippers	99.1	18.4	12.2	22.7	79.5	51.7	56.4	59.9	**115.2**	108.7
3	Utah	100.8	17.5	12.7	24.5	78.7	53.1	56.3	59.7	**115.1**	105.7
4	Phoenix	99.3	19.8	11.5	20.8	78.7	50.1	56.4	59.7	**114.9**	108.8
	Portland	100.8	16.0	9.9	23.0	76.3	49.5	54.0	57.7	**114.9**	113.4
6	Milwaukee	104.4	18.1	12.0	23.3	80.5	52.6	56.6	59.3	**114.7**	109.1
7	Denver	99.3	19.4	12.1	24.7	77.5	51.9	55.7	58.8	**114.2**	110.0
8	Atlanta	99.9	17.8	11.9	24.4	76.8	51.7	53.9	58.1	**113.0**	110.6
9	Dallas	99.5	17.4	11.1	21.1	77.8	49.7	55.0	58.2	**112.9**	110.4

Source: http://www.espn.com/nba/hollinger/teamstats

Expected (Shot) Values

How Much is a Shot Worth?

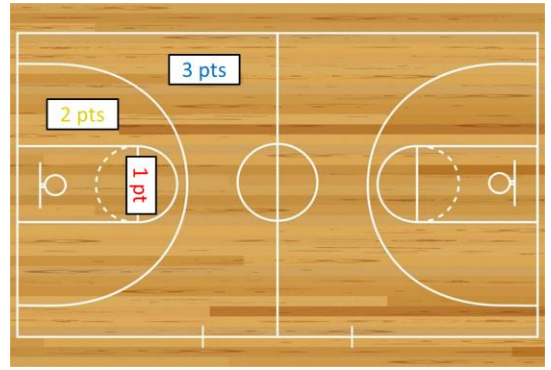
- Not all shots are created equal (or worth the same points)



Source: https://www.amazon.com/Leowefowa-Basketball-Backdrop-Background-Birthday/dp/B07YF3THL7

Effective Field Goal Percentage + True Shooting Percentage

- **Field Goal Percentage (FG%)** = $\frac{Field\ Goals\ Made\ (FGM)}{Field\ Goal\ Attempts\ (FGA)}$
  - Problems?



- Improvements:
  - **Effective Field Goal Percentage (eFG%)** = $\frac{FGM + 0.5 * 3-pointers\ made\ (3PM)}{FGA}$
  - **True Shooting Percentage (TS%)** = $\frac{PTS/_2}{FGA + 0.44 * FTA}$

How Much is a Shot Worth…Really?

- At the time of this photo, how much is this shot worth? What is its "**value**?"

Source: https://www.nytimes.com/2019/05/12/sports/nba-playoffs-trail-blazers-nuggets-sixers-raptors.html

Expected Value

- E(Something) = **Expected Value** of something
  - Fancy talk for **average**

  - Consider a binary event X, like you get a 1-point exam question right (X = 1) or wrong (X = 0). What's the expected value of that exam question?

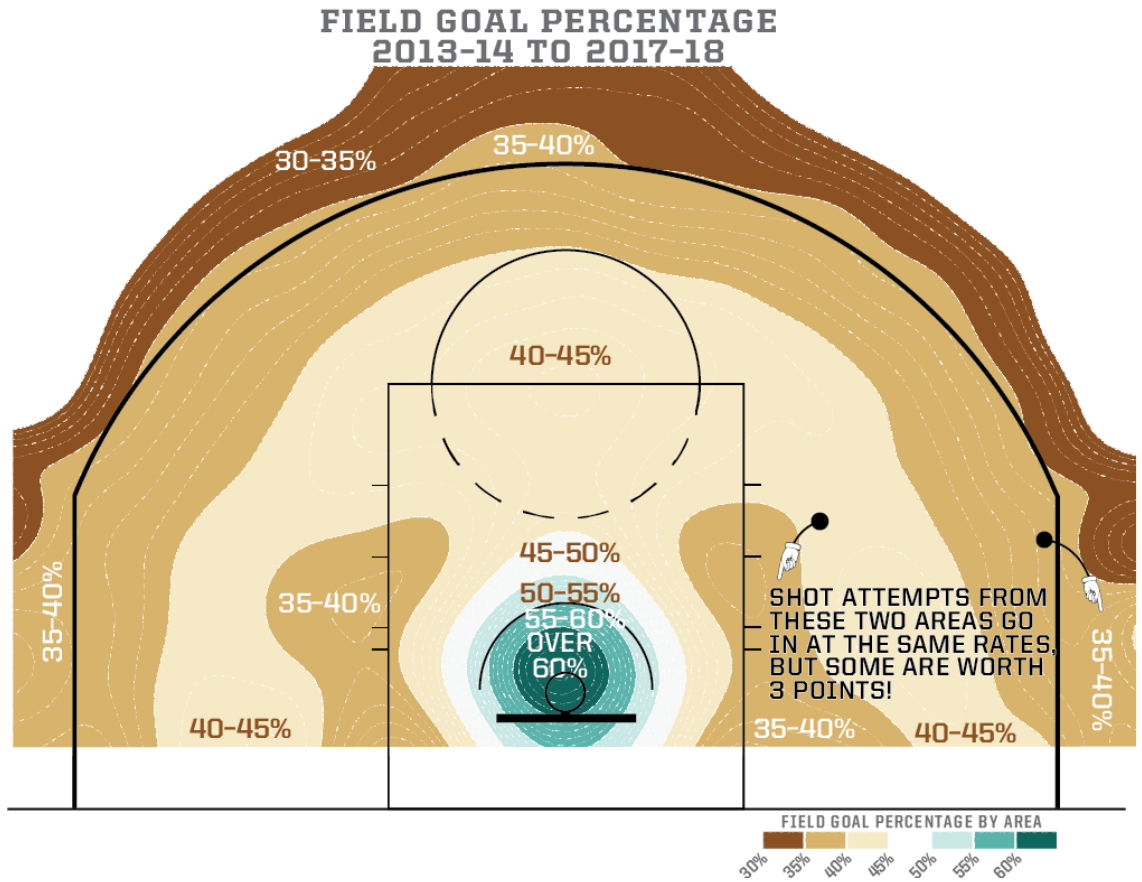
  - $E(X) = \Pr(X = 1) * 1 + \Pr(X = 0) * 0$
    - $\Pr(X = 1)$ is the probability you get the question right

Expected Value

- E(Something) = **Expected Value** of something
  - Extend this to a 3-point basketball shot
  - $E(shot) = \Pr(Made) * 3 + \Pr(Miss) * 0$
    - **Expected** value of shot (in **points**)
    - Range?
  - Can do this calculation for any and all shots – just need to know their values (0 vs. not-zero points) and probability of make vs. miss

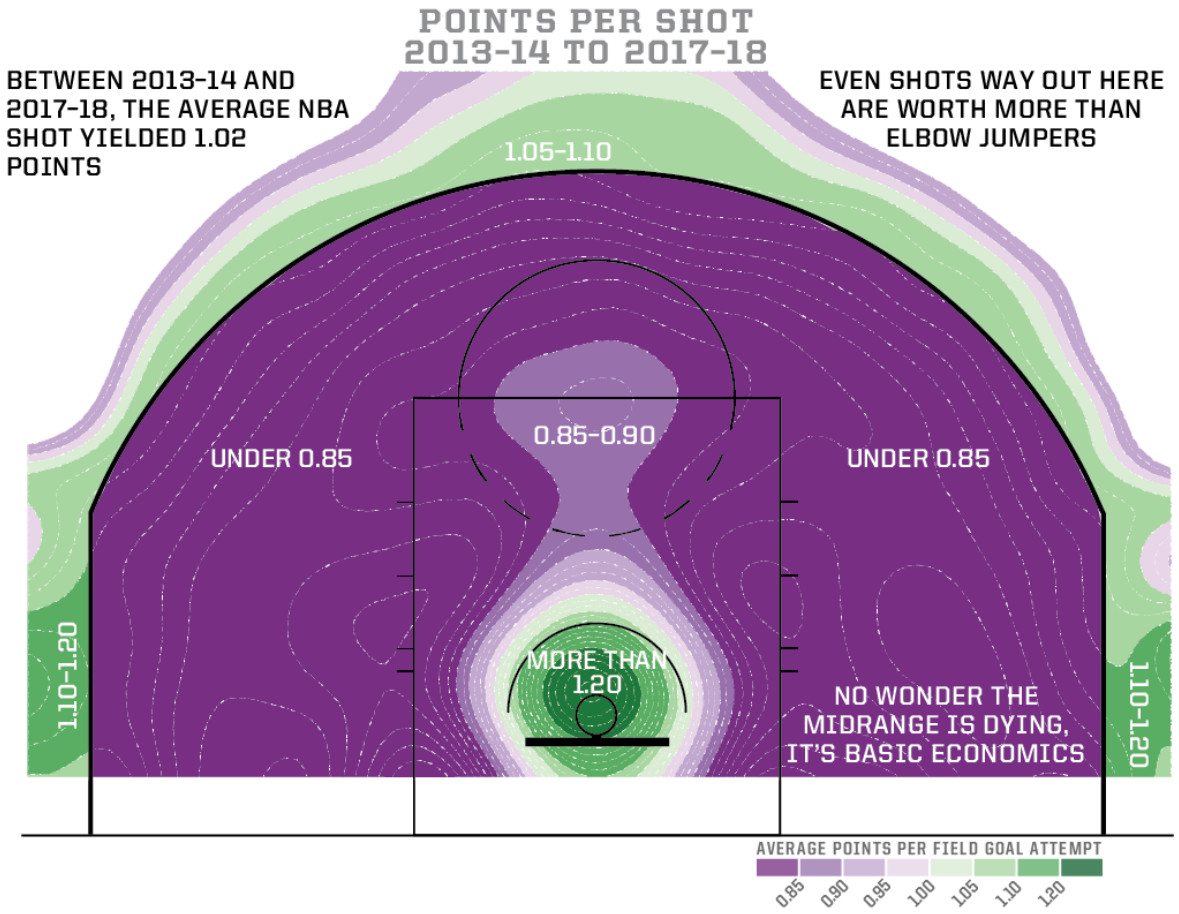
Source: https://www.nytimes.com/2019/05/12/sports/nba-playoffs-trail-blazers-nuggets-sixers-raptors.html

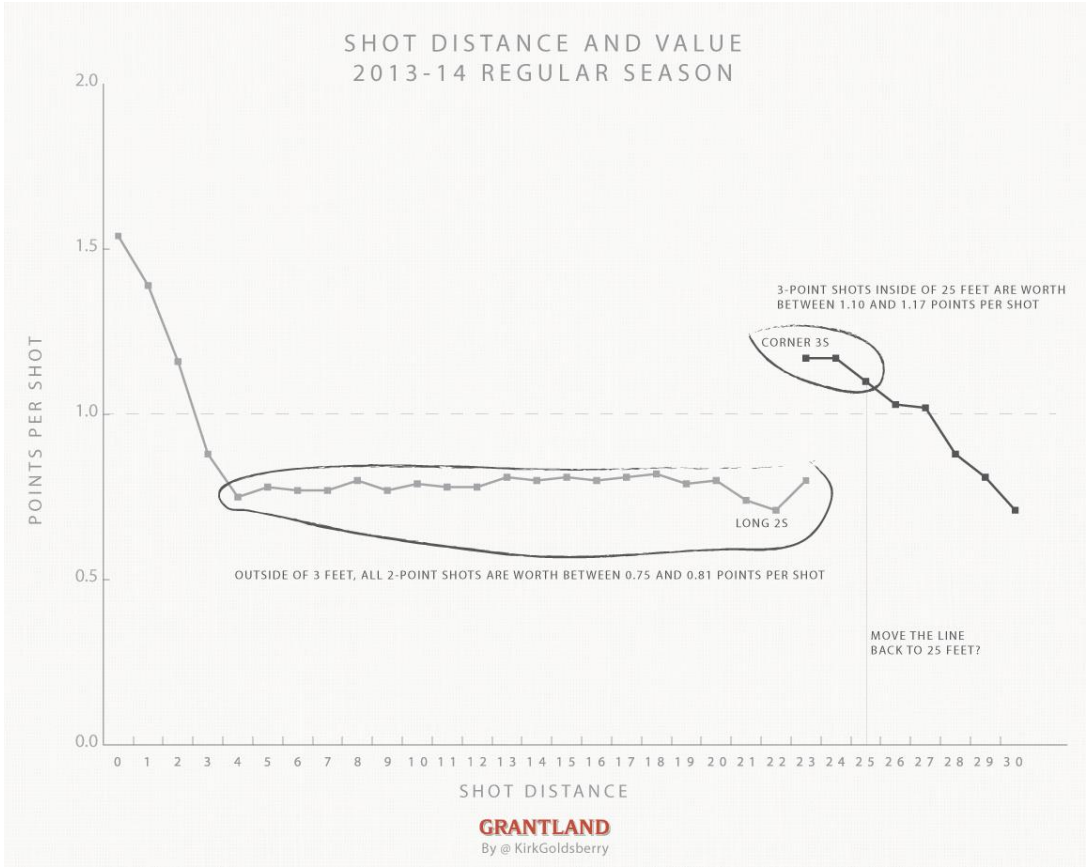
NBA Shot Value



Source: https://fivethirtyeight.com/features/how-mapping-shots-in-the-nba-changed-it-forever/

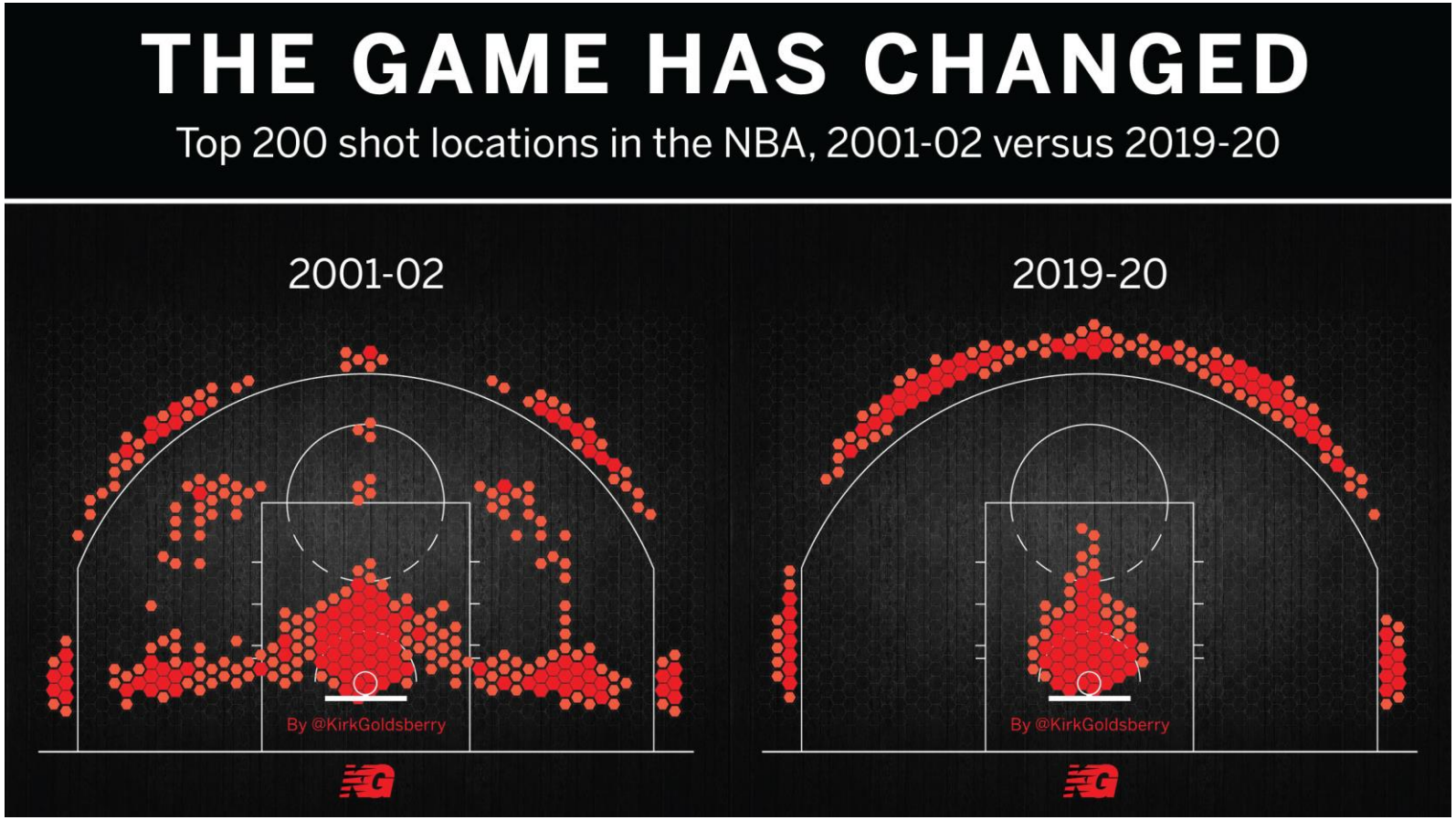
NBA Shot Value





Source: https://grantland.com/the-triangle/is-it-time-to-move-the-nba-3-point-line-back/

“Death of the Midrange Jumper”/ “Moreyball”



Source: https://twitter.com/kirkgoldsberry/status/1217109175894831105?s=20

The Quest for Holistic Statistics

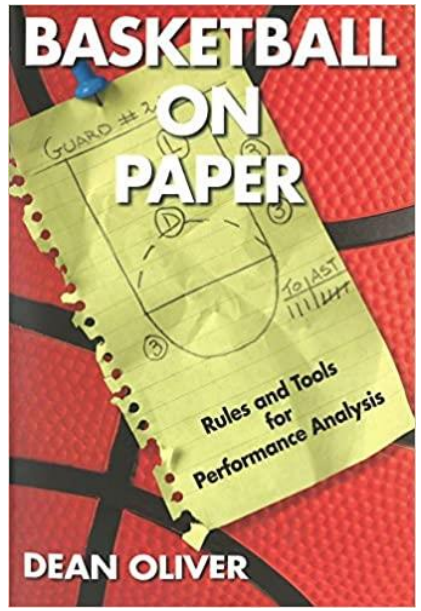
FOUR FACTORS

PER

PLUS-MINUS STATISTICS

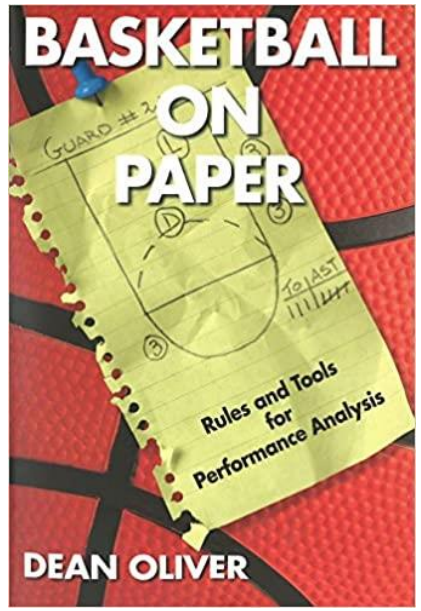
Four Factors

- Shooting isn't all there is to winning basketball games
- Dean Oliver's **Four Factors** from *Basketball on Paper:*
  - 1. Shooting/Scoring Efficiently (40%)
    - Measured with eFG%



Four Factors

- Shooting isn't all there is to winning basketball games
- Dean Oliver's **Four Factors** from *Basketball on Paper:*
  - 1. Shooting/Scoring Efficiently (40%)
  - 2. Protect the Ball/Avoid Turnovers (25%)
    - Measured with **Turnover Percentage (TOV%)** = $$\frac{TOV}{FGA + 0.44 * FTA + TOV}$$

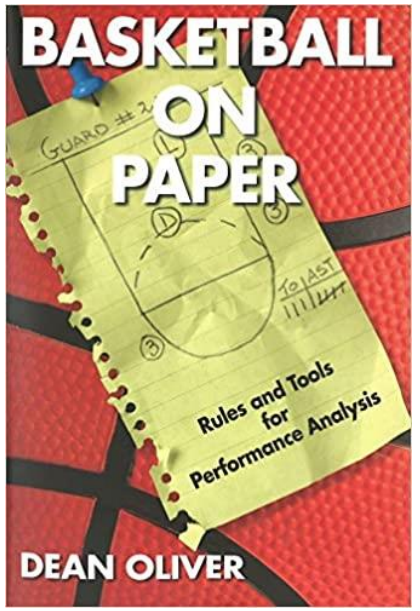


Four Factors

- Shooting isn't all there is to winning basketball games
- Dean Oliver's **Four Factors** from *Basketball on Paper:*
  - 1. Shooting/Scoring Efficiently (40%)
  - 2. Protect the Ball/Avoid Turnovers (25%)
  - 3. Rebounding (20%)
    - Measured with

      Offensive Rebounding Percentage (ORB%) = $\frac{Offensive\ Rebounds\ (ORB)}{ORB+Opp\ Defensive\ Rebounds\ (DRB)}$ *and*

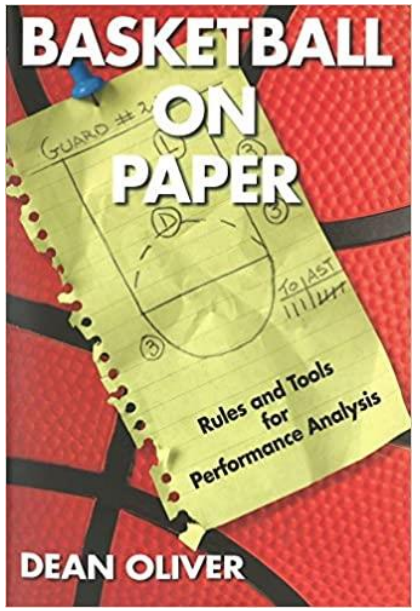
      Defensive Rebounding Percentage (DRB%) = $\frac{DRB}{DRB+Opp\ ORB}$



Four Factors

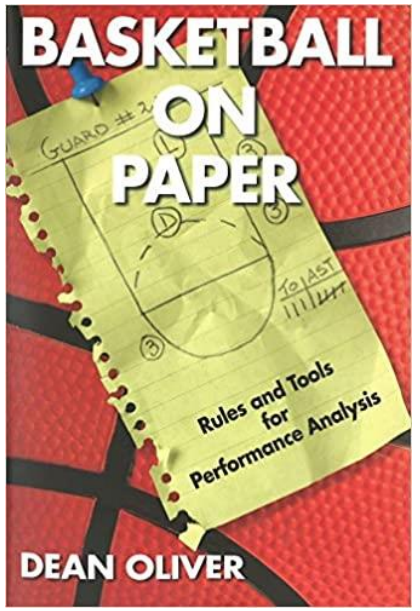
- Shooting isn't all there is to winning basketball games
- Dean Oliver's **Four Factors** from *Basketball on Paper:*
  - 1. Shooting/Scoring Efficiently (40%)
  - 2. Protect the Ball/Avoid Turnovers (25%)
  - 3. Rebounding (20%)
  - 4. Make Lots of Free Throws (15%)
    - Measured with

      Free throw rate (FTR) = $\frac{Made\ Free\ Throws\ (FT)}{FGA}$



Four Factors

- Dean Oliver's **Four Factors** from *Basketball on Paper:*
  - 1. Shooting/Scoring Efficiently (40%)
  - 2. Protect the Ball/Avoid Turnovers (25%)
  - 3. Rebounding (20%)
  - 4. Make Lots of Free Throws (15%)
- Really Eight Factors – offense + defense. But doesn't roll off the tongue, does it?
- Weights are debated; Oliver's from a program called "Roboscout" that sounds like a regression model of some kind?



Player Efficiency Rating (PER)

- **Player Efficiency Rating (PER)**, from John Hollinger: "The PER sums up all a player's positive accomplishments, subtracts the negative accomplishments, and returns a per-minute rating of a player's performance."
- Formula = complicated, but incorporates 2- and 3-point shots, free throws, assists, steals, blocks, turnovers, fouls, offensive + defensive rebounds
- Underrates defense?

2020-21 Hollinger NBA Player Statistics - All Players

Season: 2020-21 Regular Season

Team Statistics »

League: NBA | East | West

Position: All | PG | SG | SF | PF | C | Rookies | International

Qualified* | All Players

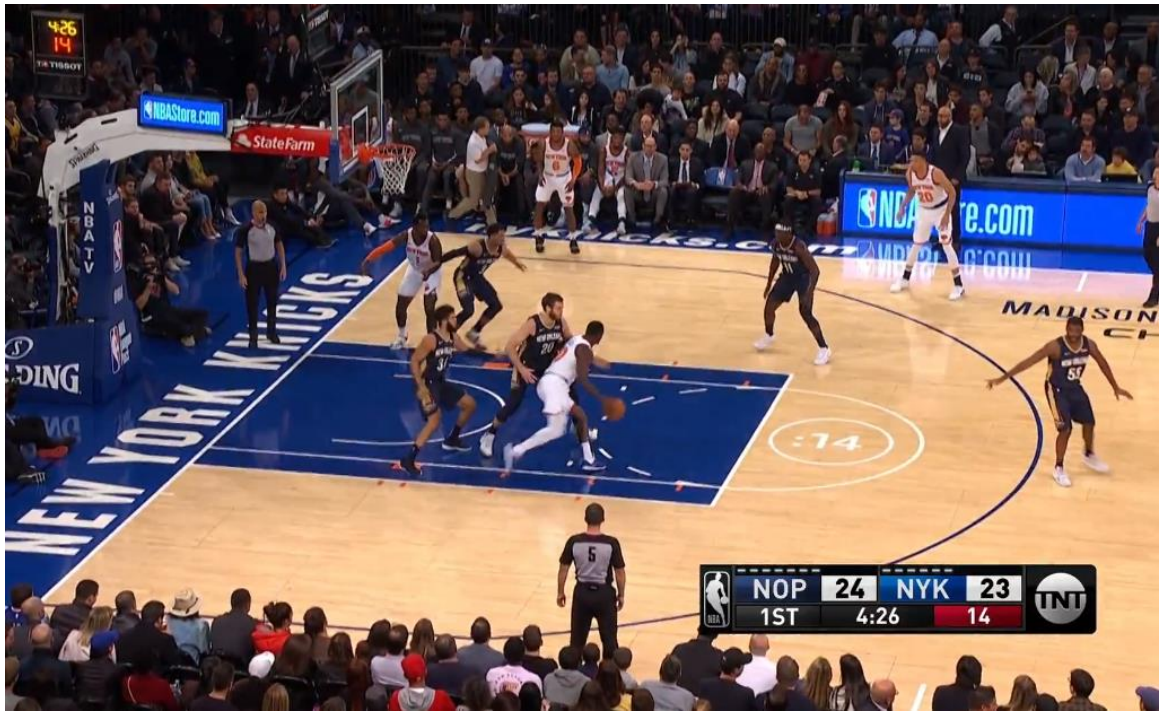
Hollinger Stats - Player Efficiency Rating - Qualified Players

RK	PLAYER	GP	MPG	TS%	AST	TO	USG	ORR	DRR	REBR	PER	VA	EWA
1	Nikola Jokic, DEN	72	34.6	.647	26.2	9.7	35.9	7.9	22.1	15.1	**31.36**	770.8	25.7
2	Joel Embiid, PHI	51	31.1	.636	10.0	11.0	39.4	6.7	25.6	16.4	**30.32**	466.4	15.5
3	Giannis Antetokounmpo, MIL	61	33.0	.633	18.6	10.8	39.1	4.8	24.2	14.8	**29.24**	0.0	0.0
4	Zion Williamson, NO	61	33.2	.649	13.6	10.0	34.2	7.4	12.5	10.0	**27.17**	0.0	0.0
5	Jimmy Butler, MIA	52	33.6	.607	26.4	7.8	32.5	5.8	14.7	10.4	**26.57**	0.0	0.0
6	Kevin Durant, BKN	35	33.1	.666	19.1	11.7	36.2	1.2	18.3	10.2	**26.44**	0.0	0.0
7	Stephen Curry, GS	63	34.2	.655	17.2	10.1	37.9	1.5	13.3	7.4	**26.37**	0.0	0.0
8	Kawhi Leonard, LAC	52	34.1	.622	19.1	7.4	34.4	3.3	14.7	9.1	**26.09**	0.0	0.0
9	Robert Williams III, BOS	52	18.9	.719	21.6	11.9	18.4	13.3	22.4	17.8	**25.71**	222.1	7.4
10	Damian Lillard, POR	67	35.8	.623	22.4	9.0	39.1	1.3	9.8	5.4	**25.65**	0.0	0.0

Source: http://insider.espn.com/nba/hollinger/statistics

Basketball is Complex

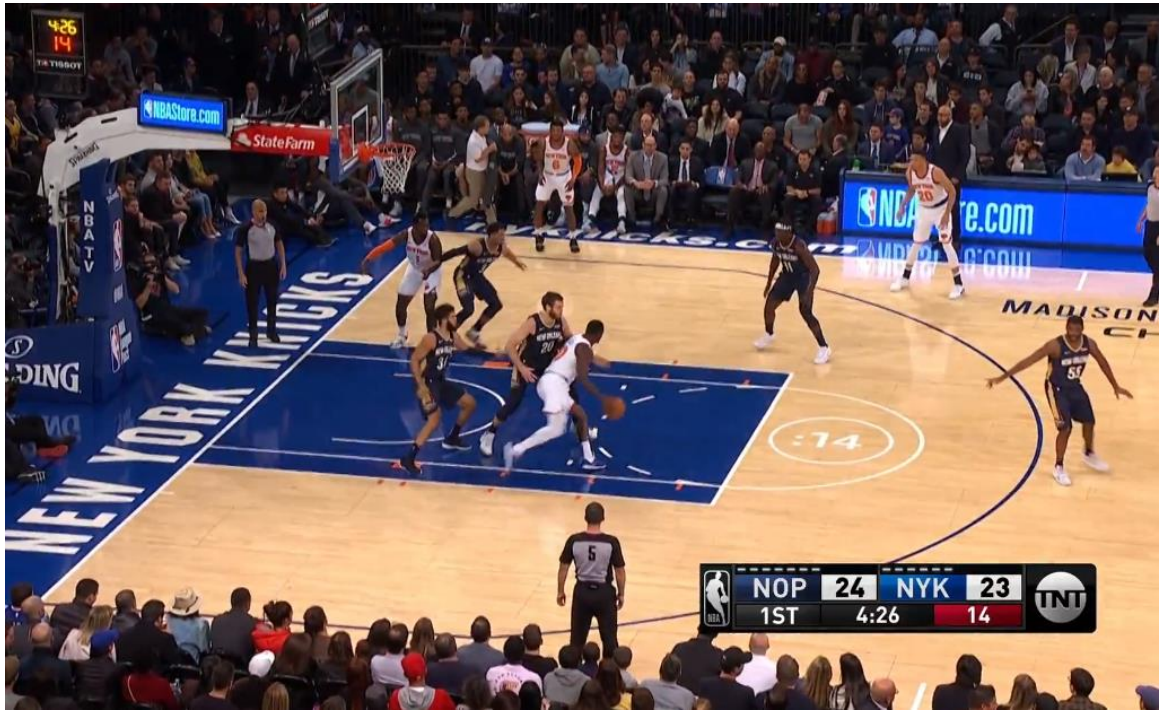
- What's the point of basketball? How do you win?
- So how might you go about measuring the value of a player and all they do (esp. defense)?



Source: https://awfulannouncing.com/turner/tnt-tries-out-an-on-court-shot-clock-graphic-to-mixed-reviews.html; https://tht.fangraphs.com/can-we-predict-hitters-change-based-on-how-pitchers-approach-them/

Basketball is Complex

- How do you go about looking at the value of anything, like…oh, I dunno, let's say just for no particular reason…a vaccine?



Source: https://awfulannouncing.com/turner/tnt-tries-out-an-on-court-shot-clock-graphic-to-mixed-reviews.html; https://tht.fangraphs.com/can-we-predict-hitters-change-based-on-how-pitchers-approach-them/

Plus-Minus

- **(Traditional) Plus-Minus (+/-)**: score differential when player is in the game

- Problems?

Hawks

STARTERS	MIN	FG	3PT	FT	OREB	DREB	REB	AST	STL	BLK	TO	PF	+/-	PTS
J. Collins PF	41	5-6	0-1	4-4	3	13	16	1	1	1	2	4	+7	14
C. Capela C	32	6-8	0-0	1-4	0	6	6	1	1	0	0	3	-2	13
T. Young PG	43	5-23	2-11	9-11	0	3	3	10	0	0	6	4	+7	21
B. Bogdanovic SG	21	2-8	0-4	0-0	0	2	2	2	2	0	0	1	-8	4
K. Huerter SG	40	10-18	2-4	5-7	1	6	7	3	1	0	1	4	+8	27
BENCH	MIN	FG	3PT	FT	OREB	DREB	REB	AST	STL	BLK	TO	PF	+/-	PTS
D. Gallinari PF	30	6-13	3-7	2-2	1	4	5	2	1	0	1	2	+8	17
O. Okongwu PF	16	0-1	0-0	1-2	1	2	3	1	0	0	0	3	+9	1
S. Hill SF	5	0-0	0-0	0-0	0	0	0	0	0	0	0	0	+5	0
L. Williams SG	11	2-4	0-0	2-2	0	2	2	2	3	0	0	2	+1	6

76ers

STARTERS	MIN	FG	3PT	FT	OREB	DREB	REB	AST	STL	BLK	TO	PF	+/-	PTS
T. Harris PF	45	8-24	2-7	6-6	2	12	14	4	2	0	2	4	-6	24
J. Embiid C	41	11-21	2-5	7-10	1	10	11	3	1	1	8	3	-2	31
B. Simmons PG	36	2-4	0-0	1-2	1	7	8	13	1	0	2	4	+1	5
S. Curry SG	31	6-10	3-5	1-2	0	2	2	2	1	1	1	5	-2	16
F. Korkmaz SG	18	2-6	1-5	0-0	0	2	2	1	0	0	0	3	+5	5
BENCH	MIN	FG	3PT	FT	OREB	DREB	REB	AST	STL	BLK	TO	PF	+/-	PTS
D. Howard C	7	0-2	0-0	0-0	1	2	3	0	0	1	1	3	-5	0
T. Maxey PG	14	0-2	0-1	2-2	0	1	1	3	0	0	2	2	-7	2

Source: https://www.espn.com/nba/boxscore?gameId=401332958

On-Off Court Splits

- **On-Off splits:** addresses problem of how team does with *compared to without* this player
- Problems?

JOEL EMBIID

7'0" BIG 27.4 YEARS OLD #3 PICK IN 2014

CONTRACT: $29.5M / $31.6M / $33.6M → TEAM SALARY SHEET

PROFILE POSITIONS ON/OFF STATS GAME LOGS LINEUPS TRENDS

Team Efficiency and Four Factors

The "on/off diff" rows show the **difference** in how the team performed with the player on vs. off the court for each of these stats.
The orange/blue numbers show the player's percentile rank in that stat relative to all players.

Show On-Court Stats

					OFFENSE												DEFENSE									
	Year	Age	Team	MIN	Diff		Exp W	Pts/Poss		eFG%		TOV%		ORB%		FT Rate		Pts/Poss		eFG%		TOV%		ORB%		FT Rate	
On/Off Diff	16-17	22.6	PHI	781	96	+12.7	+32	69	+1.9	70	+0.9%	15	+1.5%	73	+1.6%	100	+9.6	99	-10.8	97	-4.6%	48	-0.1%	86	-2.4%	78	-2.6
On/Off Diff	17-18	23.6	PHI	1899	97	+13.0	+30	89	+5.6	79	+1.6%	87	-1.4%	37	-1.1%	99	+8.5	94	-7.4	95	-3.7%	8	-2.0%	51	-0.1%	91	-3.9
On/Off Diff	18-19	24.6	PHI	2156	98	+14.4	+35	93	+6.8	91	+2.8%	57	-0.2%	29	-1.4%	100	+9.3	96	-7.6	91	-2.9%	30	-0.7%	81	-1.9%	91	-3.6
On/Off Diff	19-20	25.6	PHI	1503	82	+5.8	+15	28	-3.3	16	-2.6%	19	+1.4%	75	+2.1%	100	+8.8	96	-9.0	95	-4.4%	24	-0.9%	85	-2.4%	83	-2.5
On/Off Diff	20-21	26.6	PHI	1584	97	+13.0	+28	97	+10.5	99	+5.7%	83	-1.3%	7	-4.7%	100	+9.6	74	-2.5	72	-1.1%	47	-0.0%	47	+0.1%	95	-4.8

Source: https://cleaningtheglass.com/stats/player/1049/onoff#tab-team_efficiency

Defining Player Value – Question Type

- **1. Descriptive**
  - Describing the world as it *is*
- **2. Predictive**
  - Describing the world as it *will be*
- **3. Causal (Counterfactual Prediction)**
  - Describing the world as it *could be*

Plus-minus, on/off is a correlation. Distinguishing that from causation adds a whole other layer of difficulty!

Adjusted Plus-Minus

- **Adjusted Plus-Minus (APM)**: adjusts +/- for presence of other players, opponents, and various other factors (like home vs. away, coaches, etc.) depending on particular model
  - Adds context; deals with how team does when player is on with certain players, on with other players, or off
  - Uses **multivariable regression models**
  - Huh? Bear with me…

Adjusted Plus-Minus

- APM calculations, the basics:
- **Stint** = period of the game with no subs

defense. Before we get into the nitty gritty details, consider the general idea. Say you have three players and they are playing in a 2 on 2 basketball game. The plus-minus splits look like this:

P1 + P2 on the court: +10 points
P1 + P3 on the court: +8 points
P2 + P3 on the court: +4 points

Source: https://fansided.com/2014/09/25/glossary-plus-minus-adjusted-plus-minus/

Adjusted Plus-Minus

- APM calculations – the basics
- 3x3 game, 2 teams of 5 players:

+/- per 100 possessions = +6*100/15 = 40

We will identify the players as A1, A2, A3, A4, and A5 for Team A, and [illegible]2, B3, B4, and B5 for Team B. For insig suppose that each rotation comes out as follows:

- A1, A2, A3 vs. B1, B2, B3 resulted in 11-5 over 15 possessions
- A1, A2, A3 vs. B1, B2, B4 resulted in 2-4 over 8 possessions; current score 13 – 9
- A1, A4, A5 vs. B1, B4, B5 resulted in 7-1 over 10 possessions; current score 20 – 10
- A1, A4, A5 vs. B3, B4, B5 resulted in 6-6 over 17 possessions; current score 26 – 16
- A2, A3, A4 vs. B2, B3, B4 resulted in 4-7 over 10 possessions; current score 30 – 23
- A2, A3, A5 vs. B2, B3, B5 resulted in 5-7 over 16 possessions; current score 35 – 30
- A1, A2, A3 vs. B2, B4, B5 resulted in 4-5 over 8 possessions; current score 39 – 35
- A1, A2, A3 vs. B1, B4, B5 resulted in 4-7 over 10 possessions; current score 43 – 42
- A3, A4, A5 vs. B1, B4, B5 resulted in 4-0 over 8 possessions; current score 47 – 42
- A1, A3, A4 vs. B1, B2, B3 resulted in 5-2 over 10 possessions; current score 52 – 44
- A1, A4, A5 vs. B2, B3, B4 resulted in 2-9 over 14 possessions; final score 54 – 53

$$Y_1 = \beta_1 A_{1,1} + \beta_2 A_{2,1} + \beta_3 A_{3,1} + \beta_4 A_{4,1} + \beta_5 A_{5,1} + \beta_6 B_{1,1} + \beta_7 B_{2,1} + \beta_8 B_{3,1} + \beta_9 B_{4,1} + \beta_{10} B_{5,1}$$

$$Y_2 = \beta_1 A_{1,2} + \beta_2 A_{2,2} + \beta_3 A_{3,2} + \beta_4 A_{4,2} + \beta_5 A_{5,2} + \beta_6 B_{1,2} + \beta_7 B_{2,2} + \beta_8 B_{3,2} + \beta_9 B_{4,2} + \beta_{10} B_{5,2}$$

$$\vdots$$

$$Y_s = \beta_1 A_{1,s} + \beta_2 A_{2,s} + \beta_3 A_{3,s} + \beta_4 A_{4,s} + \beta_5 A_{5,s} + \beta_6 B_{1,s} + \beta_7 B_{2,s} + \beta_8 B_{3,s} + \beta_9 B_{4,s} + \beta_{10} B_{5,s}.$$

Let's define some parts of this equation. First, the quantity **Y_i** is the average point differential of home team over away team per 100 possessions. The values **X_ij** identifies if player **i** on team **X** is playing in stint, **j**. To be complete, let's assume Team A is the home team. Then, this value is -1 for a road player, 1 for a home player, and 0 if that player is not on the court. Let's plug in all the pieces for our eleven stints:

$$\begin{aligned}
40 &= \beta_1 + \beta_2 + \beta_3 - \beta_6 - \beta_7 - \beta_8 \\
-25 &= \beta_1 + \beta_2 + \beta_3 - \beta_6 - \beta_7 - \beta_9 \\
60 &= \beta_1 + \beta_4 + \beta_5 - \beta_6 - \beta_9 - \beta_{10} \\
0 &= \beta_1 + \beta_4 + \beta_5 - \beta_8 - \beta_9 - \beta_{10} \\
-30 &= \beta_2 + \beta_3 + \beta_4 - \beta_7 - \beta_8 - \beta_9 \\
-12.5 &= \beta_2 + \beta_3 + \beta_5 - \beta_7 - \beta_8 - \beta_{10} \\
-12.5 &= \beta_1 + \beta_2 + \beta_3 - \beta_7 - \beta_9 - \beta_{10} \\
-30 &= \beta_1 + \beta_2 + \beta_3 - \beta_6 - \beta_9 - \beta_{10} \\
50 &= \beta_3 + \beta_4 + \beta_5 - \beta_6 - \beta_9 - \beta_{10} \\
30 &= \beta_1 + \beta_3 + \beta_4 - \beta_6 - \beta_7 - \beta_8 \\
-50 &= \beta_1 + \beta_4 + \beta_5 - \beta_7 - \beta_8 - \beta_9
\end{aligned}$$

Source: https://squared2020.com/2017/09/18/deep-dive-on-regularized-adjusted-plus-minus-i-introductory-example/

Adjusted Plus-Minus

- **Adjusted Plus-Minus (APM)**: adjusts +/- for presence of other players, opponents, and various other factors (like home vs. away, coaches, etc.) depending on particular model
  - Remaining problems?
    - 1. Collinearity – some players play together a *lot*, others barely (or not at all). Use "Regularization" to deal with that → **Regularized Adjusted Plus-Minus (RAPM)**
      - Better predictive ability for future years
    - 2. APM varies wildly year-to-year (not "repeatable"), depends on new players, coaching and role changes, etc.
    - 3. *Why* is a player good?
      - Implications for team-building?
    - 4. Nick Collison Effect – fixes?

Source: https://fansided.com/2014/09/25/glossary-plus-minus-adjusted-plus-minus/

Lineup Stats

- Due to limitations of individual +/-, just look at holistic lineup stats?

SEASON: 2020-21 | SEASON TYPE: Regular Season | TEAM: All Teams | LINEUPS: 5 Player Lineups | Advanced Filters

RECENT FILTERS | GLOSSARY | SHARE

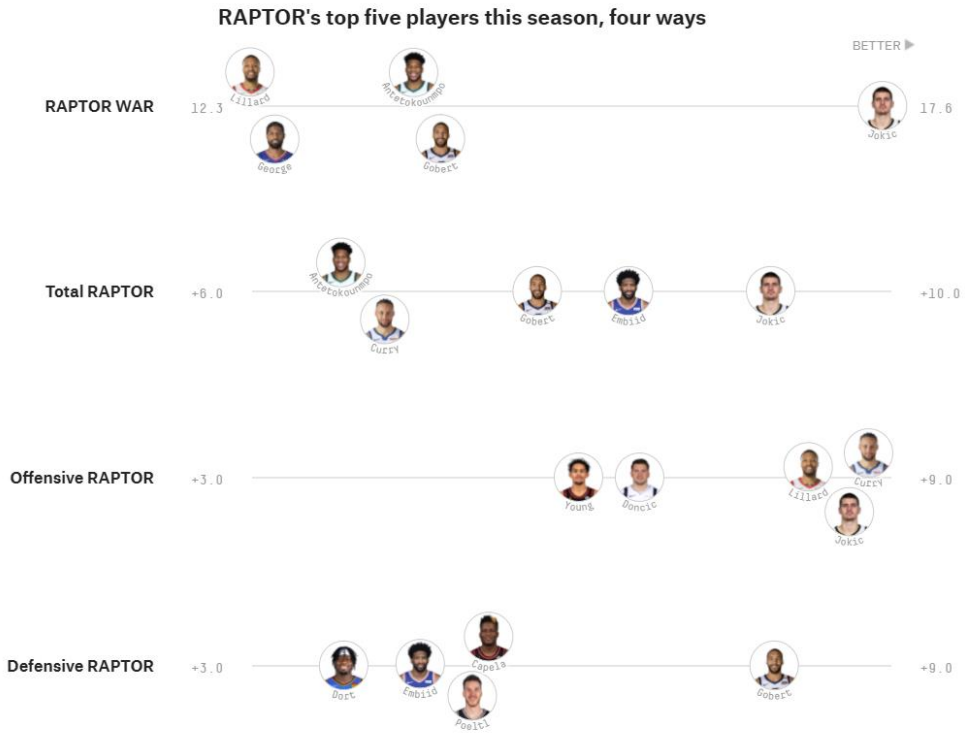
2000 Rows | Page 1 of 80

LINEUPS	TEAM	GP	MIN	OFFRTG	DEFRTG	NETRTG	AST%	AST/TO	AST RATIO	OREB%	DREB%	REB%	TO RATIO	EFG%	TS%	PACE	PIE
.C. Paul, .J. Crowder, .D. Booker, .M. Bridges, .D. Ayton	PHX	52	706	115.9	111.2	4.7	63.3	2.42	20.3	21.7	74.8	49.0	0.1	57.9	61.1	100.11	53.7
.D. Green, .T. Harris, .S. Curry, .J. Embiid, .B. Simmons	PHI	32	656	117.7	103.7	14.0	60.9	2.02	19.0	21.7	76.0	50.7	0.1	57.9	62.1	100.57	58.9
.E. Bledsoe, .S. Adams, .B. Ingram, .L. Ball, .Z. Williamson	NOP	41	610	115.9	113.8	2.0	64.4	2.17	20.4	30.0	75.5	53.2	0.1	57.1	59.0	100.35	52.8
.M. Conley, .B. Bogdanovic, .R. Gobert, .R. O'Neale, .D. Mitchell	UTA	41	593	114.2	103.4	10.8	55.7	1.70	16.6	27.6	73.6	52.4	0.1	55.8	59.2	99.89	55.8
.N. Noel, .R. Bullock, .E. Payton, .J. Randle, .R. Barrett	NYK	41	554	109.0	113.6	-4.6	55.0	1.65	16.8	24.6	73.4	48.9	0.1	54.3	57.0	97.35	45.4
.B. Lopez, .J. Holiday, .K. Middleton, .G. Antetokounmpo, .D. DiVincenzo	MIL	45	508	117.7	109.0	8.7	60.5	2.07	19.9	23.0	80.6	52.6	0.1	59.8	61.9	107.90	55.9
.J. Valanciunas, .K. Anderson, .D. Brooks, .G. Allen, .J. Morant	MEM	35	457	116.2	107.6	8.7	60.4	2.28	19.0	26.8	78.7	53.2	0.1	55.0	58.2	104.63	53.6

Source: https://fansided.com/2014/09/25/glossary-plus-minus-adjusted-plus-minus/

Variations on a Theme

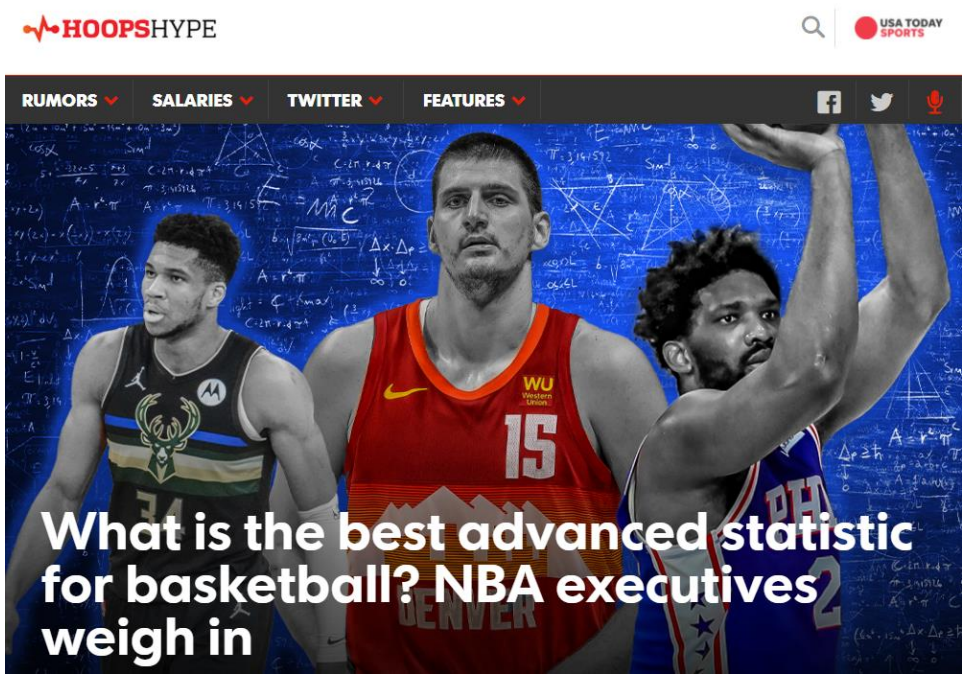
- **+/-**
- **APM**
- **RAPM**
- **Real Plus-Minus (RPM, ESPN's main stat)** – adds in some other adjustments to RAPM like aging curves and game condition
- FiveThirtyEight's CARMELO (old) and RAPTOR (current) NBA projection and player valuation systems
- LEBRON from Basketball Index, Estimated Plus-Minus, Daily Plus-Minus (DPM from DARKO)
- And so many more…



Source: https://fivethirtyeight.com/features/introducing-raptor-our-new-metric-for-the-modern-nba/

Overview of Holistic Metrics

- Great survey and overview of +/- style metrics and how NBA executives think about them



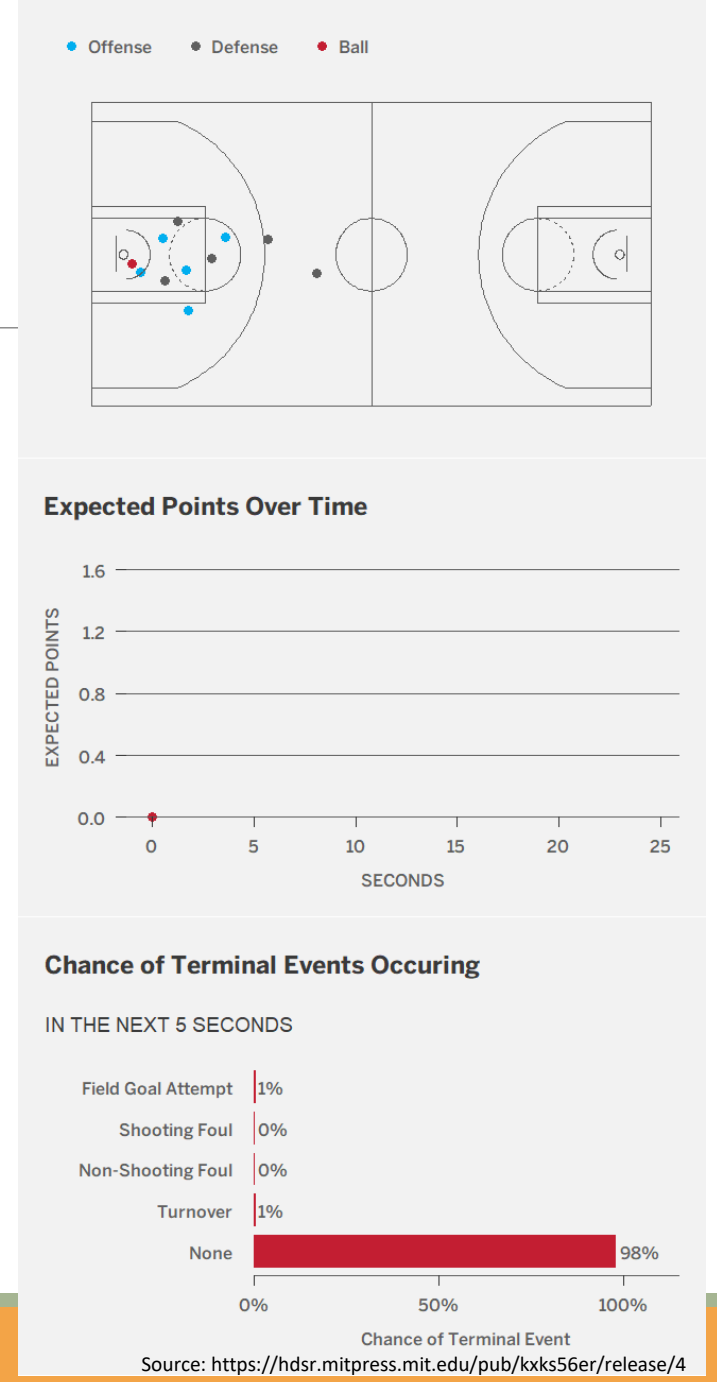
Source: https://hoopshype.com/lists/advanced-stats-nba-real-plus-minus-rapm-win-shares-analytics/

The Future: Expected Points Value Added (EPVA)

Tracking Data: Expected Points Value Added (EPVA)

- Use tracking data to estimate real-time value of a possession

- Isolate effects of individual events, then sum up all a player's actions (vs. average player)



Source: https://hdsr.mitpress.mit.edu/pub/kxks56er/release/4

Tracking Data: Expected Points Value Added (EPVA)

- Spurs-Cavs, February 13, 2013

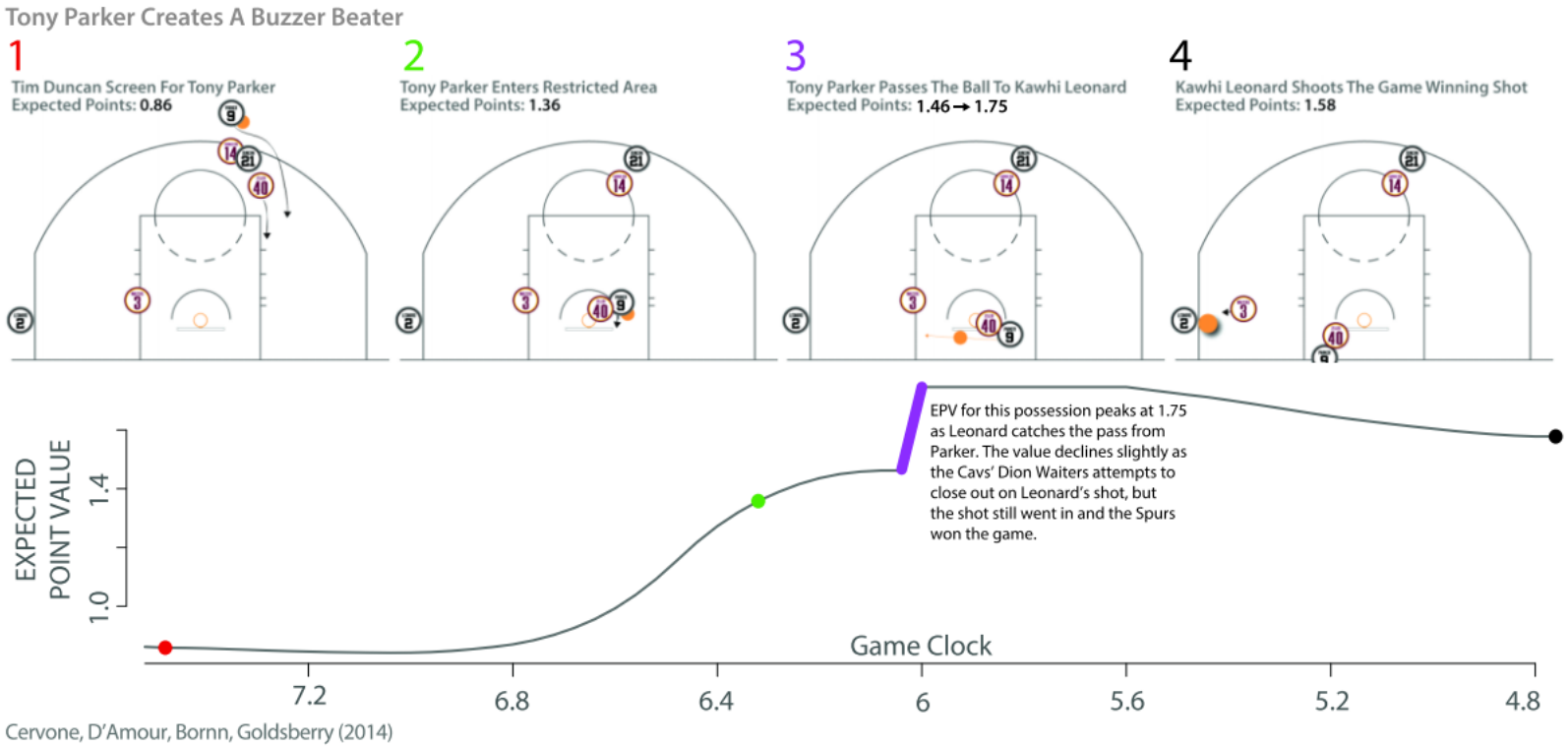


Figure 2. EPV throughout the Spurs' final possession, with annotations of major events.

Source: https://deadspin.com/kirk-goldsberrys-new-nba-stat-is-a-huge-step-for-baske-1517481614

Hot Hand Fallacy (?) and Selection Effects

The Hot Hand

Source: NBA Jam, NFL Blitz

The Hot Hand

- People are streaky
- Success breeds success
- Example: basketball players more likely to make shots after previous makes; “get in a rhythm”
  - Many other sports and skill-based activities
  - Coaches, players, experts *strongly* believe it exists

Source: https://slate.com/technology/2015/10/hot-hands-in-basketball-are-real-new-analysis-shows.html

The Hot Hand Fallacy

- Gilovich, Vallone, Tversky (GVT, 1985): ask Cornell basketball players to make 100 shots from a range where they shoot 50%

  - Also looked at '76ers shooting, Celtics paired FTs → similar results

  - People see patterns in randomness + confirmation bias

TABLE 4

Probability of Making a Shot Conditioned on the Outcome of Previous Shots for All Cornell Players

Player	*P*(hit/3 misses)	*P*(hit/2 misses)	*P*(hit/1 miss)	*P*(hit)	*P*(hit/1 hit)	*P*(hit/2 hits)	*P*(hit/3 hits)
Males							
1	.44 (9)	.50 (18)	.61 (46)	.54 (100)	.49 (53)	.48 (25)	.50 (12)
2	.43 (28)	.33 (42)	.35 (65)	.35 (100)	.35 (34)	.25 (12)	.00 (3)
3	.67 (6)	.68 (19)	.49 (39)	.60 (100)	.67 (60)	.62 (40)	.60 (25)
4	.47 (15)	.45 (29)	.43 (53)	.40 (90)	.36 (36)	.23 (13)	.33 (3)
5	.75 (12)	.60 (30)	.47 (57)	.42 (100)	.36 (42)	.40 (15)	.33 (6)
6	.25 (12)	.38 (21)	.48 (42)	.57 (100)	.65 (57)	.62 (37)	.65 (23)
7	.29 (7)	.50 (16)	.47 (32)	.56 (75)	.64 (42)	.63 (27)	.65 (17)
8	.50 (6)	.50 (12)	.52 (25)	.50 (50)	.46 (24)	.64 (11)	.57 (7)
9	.35 (20)	.33 (30)	.35 (46)	.54 (100)	.72 (53)	.79 (38)	.83 (30)
10	.57 (7)	.50 (14)	.64 (39)	.59 (100)	.79 (38)	.60 (35)	.57 (21)
11	.57 (7)	.61 (18)	.56 (41)	.58 (100)	.59 (58)	.62 (34)	.62 (21)
12	.41 (17)	.43 (30)	.46 (56)	.44 (100)	.42 (43)	.39 (18)	.43 (7)
13	.40 (5)	.62 (13)	.67 (39)	.61 (100)	.58 (60)	.56 (34)	.50 (18)
14	.50 (6)	.62 (16)	.60 (40)	.59 (100)	.58 (59)	.59 (34)	.60 (20)
Females							
1	.67 (9)	.61 (23)	.55 (51)	.48 (100)	.42 (48)	.45 (20)	.33 (9)
2	.43 (28)	.36 (44)	.31 (65)	.34 (100)	.41 (34)	.36 (14)	.40 (5)
3	.36 (25)	.38 (40)	.33 (60)	.39 (100)	.49 (39)	.42 (19)	.50 (8)
4	.27 (30)	.33 (45)	.34 (68)	.33 (100)	.29 (31)	.33 (9)	.33 (3)
5	.22 (27)	.36 (42)	.34 (64)	.35 (100)	.37 (35)	.50 (12)	.20 (5)
6	.54 (11)	.58 (26)	.52 (54)	.46 (100)	.38 (45)	.41 (17)	.29 (7)
7	.32 (25)	.28 (36)	.36 (58)	.41 (100)	.49 (41)	.65 (20)	.62 (13)
8	.67 (9)	.55 (20)	.57 (47)	.53 (100)	.50 (52)	.58 (26)	.73 (15)
9	.46 (13)	.55 (29)	.47 (55)	.45 (100)	.41 (44)	.47 (17)	.50 (8)
10	.32 (19)	.34 (29)	.46 (54)	.47 (100)	.47 (45)	.67 (21)	.71 (14)
11	.50 (10)	.56 (23)	.51 (47)	.53 (100)	.56 (52)	.50 (28)	.39 (13)
12	.32 (37)	.32 (54)	.27 (74)	.25 (100)	.20 (25)	.00 (5)	— (0)
M =	.45	.47	.47	.47	.48	.49	.49

307

Source: GVT, *Cognitive Psychology*, 17:3, 1985

The Hot Hand (Fallacy?)

- Miller and Sanjurjo (2015): found a bias lurking in these types of analyses that had gone unrecognized for *30 years*
  - Ready for your head to explode?
  - Really?
  - Everyone reach under your seats and put on your ponchos just in case

Source: https://giphy.com/explore/head-explode

The Hot Hand (Fallacy?)

- Miller and Sanjurjo (2015): found a bias lurking in these types of analyses that had gone unrecognized for *30 years*
  - Consider a fair coin flipped 100 times
    - "Fair" means P(H) = P(T) = 0.5
  - What is the probability any 1 flip is heads? P(H) = ?
  - What is the probability any 1 flip *following at least 3 heads* is heads? P(H|_HHH) = ?

Source: https://www.scientificamerican.com/article/momentum-isnt-magic-vindicating-the-hot-hand-with-the-mathematics-of-streaks/

The Hot Hand (Fallacy?)

- Miller and Sanjurjo (2015): found a bias lurking in these types of analyses that had gone unrecognized for *30 years*
  - Random sequence of 100 makes and misses, 50% chance of each (51 makes, 49 misses)
    - Of shots after 3+ makes, only 4/11 (36%) are makes
    - Repeat 1,000 times → ~46% Makes (which, you'll note, is <50%)
    - **Issue:** every streak *has to end with a miss*. Sometimes it's a Miss after 0 more Makes, sometimes after 1, 2, …
      - *Every* time there will be a Miss; BUT there's not always a Make to balance!
        - *Sometimes* 2+ Makes, but not as often as 0 Makes!
      - Our search *forces* a Miss >50% of the time! (Or a Make <50% of the time.)

Make	Make	Miss	Make	Miss	Make	Miss	Make	Miss	Miss
Miss	Miss	Miss	Make	Make	Miss	Make	Miss	Make	Miss
Miss	Make	Miss	Miss	Make	Miss	Miss	Make	Make	Miss
Miss	Miss	Make	Miss	Make	Make	Miss	Miss	Miss	Miss
Miss	Make	Make	Make	Miss	Make	Make	Make	Miss	Miss
Miss	Make	Make	Miss	Make	Make	Make	Make	Miss	Miss
Miss	Make	Miss	Miss	Miss	Make	Make	Make	Miss	Miss
Make	Miss	Make	Miss	Make	Make	Make	Miss	Miss	Miss
Miss	Make	Make	Make	Miss	Miss	Miss	Make	Make	Make
Make	Make	Miss	Make	Make	Miss	Make	Make	Make	Make

Source: https://www.scientificamerican.com/article/momentum-isnt-magic-vindicating-the-hot-hand-with-the-mathematics-of-streaks/

The Hot Hand (Fallacy?)

- Miller and Sanjurjo (2015): found a bias lurking in these types of analyses that had gone unrecognized for *30 years*
  - So if we see a basketball player make an exactly equal proportion of her shots after 3 Makes or 3 Misses, that is evidence FOR the Hot Hand – not against it!
  - If there were no Hot Hand (as with our simulation), shots should go in somewhat *less* frequently after 3+ Makes than 3+ Misses

Make	Make	Miss	Make	Miss	Make	Miss	Make	Miss	Miss
Miss	Miss	Miss	Make	Make	Miss	Make	Miss	Make	Miss
Miss	Make	Miss	Miss	Make	Miss	Miss	Make	Make	Miss
Miss	Miss	Make	Miss	Make	Make	Miss	Miss	Miss	Miss
Miss	Make	Make	Make	Miss	Make	Make	Make	Miss	Miss
Miss	Make	Make	Miss	Make	Make	Make	Make	Miss	Miss
Miss	Make	Miss	Miss	Miss	Make	Make	Make	Miss	Miss
Make	Miss	Make	Miss	Make	Make	Make	Miss	Miss	Miss
Miss	Make	Make	Make	Miss	Miss	Miss	Make	Make	Make
Make	Make	Miss	Make	Make	Miss	Make	Make	Make	Make

Source: https://www.scientificamerican.com/article/momentum-isnt-magic-vindicating-the-hot-hand-with-the-mathematics-of-streaks/

Thanks!

- Questions? zbinney@emory.edu, @binney_z on Twitter