To Crash or Not To Crash:
A quantitative look at the relationship between offensive rebounding and transition defense in the NBA

Jenna Wiens, Guha Balakrishnan, Joel Brooks, John Guttag
Massachusetts Institute of Technology,
Cambridge, MA, USA, 02139
Email: jwiens@mit.edu

Abstract

Immediately following a missed shot an offensive player can choose to crash the boards for an offensive rebound, get back on defense, or hold their current position. In this paper, we use optical tracking data to develop novel metrics to summarize a team's strategy immediately following a shot. We evaluate each metric using data from the 2011-2012 NBA season. Our results confirm that getting back on defense and neutralizing threats early in the possession contribute to a defensive success. However, tendencies to get back early on defense after a missed shot can reduce a team's probability of getting an offensive rebound by more than half.

1 Introduction

The American General George S. Patton had a simple philosophy of war, "When in doubt, attack." The 1st Duke of Wellington had a more nuanced view. A great general, he said, should "know when to retreat, and to dare to do it." When it comes to offensive rebounding in the NBA, some coaches take a Patton-like approach--they want their players to attack the offensive boards at every opportunity. Other coaches emphasize the importance of retreating to a strong defensive position. When the offense sends more players toward the basket when a shot is taken, that team has a better chance of securing an offensive rebound [1]. But sending too many players in to rebound might impede the ability of a team to get back quickly on defense. This paper provides steps toward understanding and quantifying the tradeoff between offensive rebounding and transition defense.

Figure 1 highlights a possession in a game between Boston and New Jersey. As soon as the ball is released the three offensive players closest to the basket decide to crash. In this instance Boston gets the offensive rebound. This clip is somewhat unusual since more often than not Boston retreats after a shot goes up, readying their defense for the next possession. The tradeoff between these two opposing strategies is complex. Current popular metrics often quoted by analysts, such as offensive rebound rate and transition points allowed offer no insight into what strategic choices influence a team's performance.

Figure 1: These frames are extracted from a game between the New Jersey Nets and the Boston Celtics form March 2nd 2012 [2]. In the frame on the left we see a field goal attempt from the Celtics. The frame on the right is taken 1 second after Paul Pierce releases the shot. From the time the shot was released three players have moved closer to the basket. Players don't have much time to decide what they will do after a shot is released, but what they do decide can have a significant impact on the outcome of the game.

In this paper, we attempt to quantify this trade-off using data from the STATS SportVU system [3], which contains data from 233 games from the 2011-2012 season (including playoffs). This dataset contains all the {x,y} positions of every player on the court and the {x,y,z} coordinates of the ball at 25 frames per second. We develop several new metrics based on the positioning of teams after a missed shot. These metrics are shown to relate to outcome (though in different ways), and may be used by teams to evaluate defensive (and offensive) strategies.

2 What Teams Gain by Crashing

Player movement immediately following the release of a shot (as in Figure 1) can have an impact on not only who gets the rebound if the shot is missed, but also the result of the next possession (if the other team gets the defensive rebound). To test this hypothesis we consider all missed field goals taken at least 15ft from the basket. Our final dataset consists of 6,521 instances of missed jump shots. For each instance we consider the {x,y} positions of every offensive player on the court in relation to the basket at the time the shot is released (the blue markers in Figure 2) and at the time the rebound is secured (the red markers in Figure 2). We define *reaction time* as the total time from when the shot is released to when the ball is rebounded. The mean reaction time in our dataset was approximately 2.22s (standard deviation of 0.65 seconds).

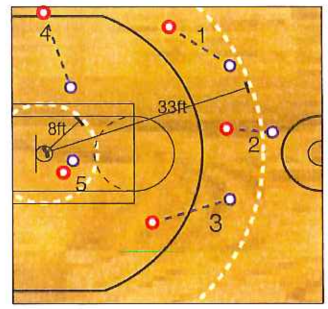


Figure 2: We consider the movement of offensive players immediately following a missed jump shot. The red markers indicate the position of players at the time the ball is released. The blue markers indicate their position at the time of the rebound.

To determine a player's intention to get the offensive rebound it is not enough to simply consider the movement of players in relation to the basket. Some players who are close to the basket may actually move slightly *away* from the basket in order to gain a better position. Players far away from the basket may not even consider crashing since the probability of getting the rebound is so low. We characterize the positioning and movement of players during the reaction time using the following four metrics:

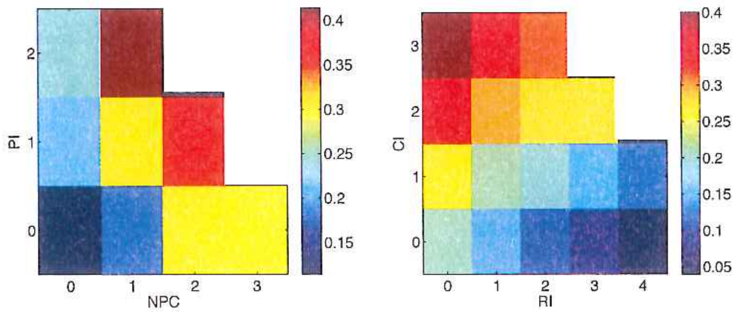
- *Number of Players Inside* (PI) – the number of players positioned 8 feet from the basket when the shot is released
- *Numbers of Players Outside* (PO) – the number of players positioned more than 33 feet from the basket when the shot is released
- *Number of Neutral Players who Crash* (NPC) –the number of players initially positioned between 8 and 33 feet from the basket who move at least 5 feet **toward** the basket
- *Number of Neutral Players who Retreat* (NPR) – the number of players initially positioned between 8 and 33 feet from the basket who move at least 5 feet **away from** the basket

In addition we define two summary statistics:

- *Crash Index* (CI) = PI+NPC
- *Retreat Index* (RI) = PO+NPR

Figure 2 is an example of Golden State's offensive positioning and movement during the reaction time after a jump shot. In this instance Golden State has PI=1, NPC=1, NPR=3 and PO=0, and was successful in securing the offensive rebound.

We begin by considering how these metrics relate to the probability of getting the offensive rebound. Considering all 6,521 possessions, we vary the NPC from 0 to 3 and the PI from 0 to 2. There were a few cases where CI=4, but these were removed since the number of instances was so low. All other combinations had at least 100 examples. Figure 3 (a) shows that the probability of getting the offensive rebound is greatest when PI=2 and NPC=1. Unsurprisingly the offensive rebounding rate increases as PI and NPC increase. Figure 3(b) shows the probability of getting the offensive rebound is greatest when no one retreats and 3 players crash. It is important to note that this figure shows an *association* between the number of players who crash and the probability of getting an offensive rebound. It does not show that crashing the boards *causes* offensive rebounds. It is possible that more players crash when it appears that getting an offensive rebound is there for the taking.



(a) (b)

Figure 3: The probability of getting an offensive rebound as a function of (a) NPC & PI and (b) CI & RI. As you increase PI and NPC the probability increases, however increasing RI always results in a lower probability.

For each team we consider all possessions where a missed jump shot occurs, and calculate the fraction of time CI>RI (more players crash than retreat), and the offensive rebounding rate. We believe that the fraction of time CI>RI is indicative of the relative importance a team places on offensive rebounding or transition defense. We plot the results of this analysis in Figure 4. Note that the SportVU system is not installed in all arenas. Therefore, there are some teams for which we have little data. To avoid confusing the signal with the noise we do not plot results for teams for which we have fewer than 150 examples of a missed jump shot. Figure 4 indicates a strong association between CI>RI and offensive rebounding.

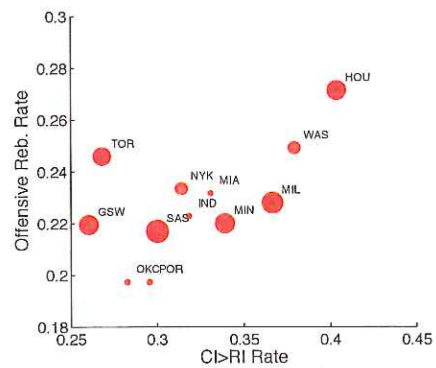


Figure 4: Considering possessions grouped by teams, we again note a relationship between CI & RI and the offensive rebounding rate (for missed jump shots in our data). The number of possessions we have for each team varies from 151 to 622, here the size of the markers indicates the number of possessions.

Moving toward the basket immediately following a missed shot seems to increase a team's probability of getting the offensive rebound, but still even the best offensive rebounding percentage is much less than 50%. I.e., following a missed jump shot, a defensive rebound by the other team is always more likely. The data shows that most of the time a team's RI is greater than its CI (42% of the time RI>CI, 32% of time CI>RI). Does sending players back early affect the result of the next defensive possession?

To address this question we consider all missed shots that result in a defensive rebound by the other team and the outcome of the next possession. We assign a 0 to possessions that result in either a made shot or an attempted free throw, and +1 to the other possessions. Here, a positive outcome is good for the team that just missed a field goal.

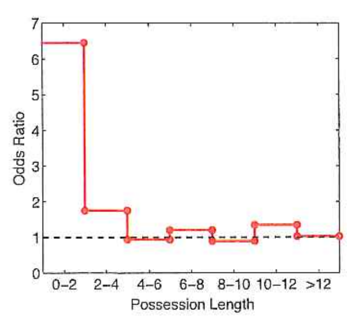


Figure 5: The positive effect of retreating immediately following the release of a shot (assuming the other team gets the defensive rebound) is transient.

For each possession we calculated the RI of the offensive team, and grouped possessions into high and low RI (upper quartile and lower quartile). We measure the probability of a positive event occurring in both groups. Figure 5 shows the impact of a team's movement immediately following the release of a shot makes in terms of outcome (assuming a missed offensive rebound). For each bin we calculate the odds ratio (Eq. 1).

$$\text{Odds Ratio} = P_{upper}(+1)^{*}P_{lower}(0)/(P_{lower}(+1)^{*}P_{upper}(0) \qquad \textbf{Eq. 1}$$

As the length of the possession increases, the odds ratio decreases. Figure 5 implies that movement away from the basket immediately following the release of the shot has the greatest impact on short possessions. I.e., the effect is transient. But what is the team giving up in return for limiting transition baskets?

Again we consider the movement of offensive players immediately following a missed shot, but in addition we consider the outcome of what happens next. If an offensive rebound occurs, any points scored during this possession count as a gain. If a defensive rebound occurs, any points scored on the next possession count as a loss. By averaging over all possessions we can compute the net gain of all missed shots (Eq. 2)

$$\text{Net Gain} = P(\text{Offensive Rebound})^{*}\text{Avg. Pts. For} - P(\text{Defensive Rebound})^{*}\text{Avg. Pts. Against} \qquad \textbf{Eq. 2}$$

Table 1 shows the average net gain (points/possession) as a function of CI and RI. In all cases, the net gain is negative, i.e., missed shots are bad. On average a missed shot results in -0.549 points per possession.

Table 1: Net gain of missed shot broken down by CI and RI. Offensive rebounding rates are given in parentheses. We consider only scenarios for which we have at least 100 possessions. The shaded area represents the most probable scenarios (over 500 instances each).

Retreat Index \ Crash Index	0	1	2	3
0	-0.83 (0.21)	-0.49 (0.27)	-0.22 (0.34)	-0.12 (0.40)
1	-0.69 (0.15)	-0.53 (0.21)	-0.45 (0.29)	-0.09 (0.36)
2	-0.76 (0.11)	-0.62 (0.19)	-0.46 (0.27)	-0.36 (0.31)
3	-0.97 (0.08)	-0.85 (0.15)	-0.40 (0.28)	
4	-1.08 (0.04)	-0.50 (0.12)		

Notice that it appears that a team can mitigate the effect of missed jump shots by sending more players to crash. Considering only the most probable scenarios (the shaded region in Table 1) we can increase the net gain from -0.62 (CI=1 & RI=2) to -0.45 (CI=2 & RI=1) just by swapping the role of a single player. This change results in a net gain of 0.17 pts/possession. If a team misses on average 25 jump shots per game this could translate to a possible gain of 4 points per game.

We investigated which teams were doing this already, by considering the ratio of the number of times teams found themselves in each scenario. We plot this ratio vs. the average net gain per team in Figure 6. Here we consider only teams for which we had at least 100 possessions (that meet our inclusion criteria) worth of data.

Admittedly this analysis does not account for the personnel on a team. For example, some teams may not have two good rebounders. In such a case sending two players to crash and one player to retreat might not have the same expected gain as it does when we consider aggregate data from all teams. This is an opportunity for further analysis.

Still our results suggest that in general focusing on the offensive rebound immediately after the shot goes up seems to trump the gain a team gets with a head start on getting back. In the next sections we investigate how a team's movement *after* the ball is rebounded impacts the outcome of a possession immediately following a missed offensive rebound.

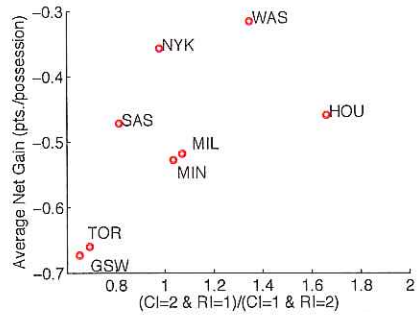
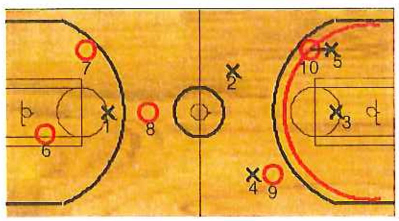
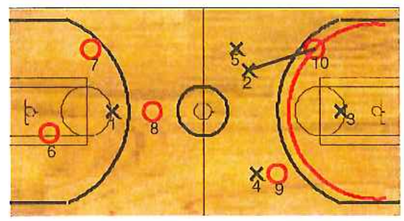


Figure 6: Washington and Houston already appear to be doing the optimal thing after a missed jump shot (at least in the games for which we had data), while Toronto and Golden State could improve their strategy.

3 Early Threat Neutralization



(a). Good threat neutralization



(b). Bad threat neutralization

Figure 7: Illustration of threat neutralization, where offense is marked with 'o' and defense is marked with 'x'. Player 10 of the offensive team has just come within his threatening distance (red arc). (a) shows an example of good neutralization, where the distance between Player 10 and the closest defender (Player 5) is small. (b) shows an example of bad neutralization.

The previous measures do not quantify how well a team prevents its opponents from entering a threatening offensive position after a defensive rebound. Simply retreating to the defensive end of the court is not sufficient. We developed the maximum distance to early threats (MDET) score to measure this element of defense. Figure 7 illustrates the intuition of the MDET score. The first time an offensive player is within a threatening distance to the basket in the first five seconds of his team's possession, we measure his distance (in feet) to the closest defending player. We compute a distance for each offensive player that becomes a threat during the possession and use the maximum of these distances as the MDET score. The maximum distance represents the most "open" player during the early portion of the possession. Note that a lower MDET score indicates better neutralization.

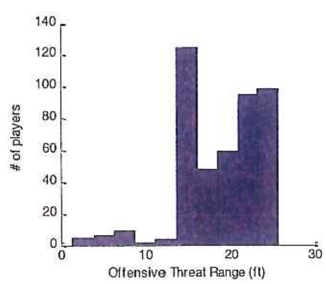


Figure 8: Distribution of offensive threat range across all players. A significant number of players are not considered a threat outside of 16 ft.

Offensive threat range varies by player. A center like Dwight Howard should be considered an offensive threat at a much closer distance to the basket than a 3-point shooter like Ray Allen. Therefore we set an offensive player's threat range based on the distribution of his shot distances in the data. We found that using the 75th percentile of a player's shot distance distribution works well. Figure 8 plots the offensive threat ranges for all the players in the data. As expected, this distribution is somewhat bimodal, with high density around 15 and 23 ft.

We compare MDET with defensive outcome by considering all continuous possessions (i.e. possessions without a stoppage in the middle) following a defensive rebound. During our analysis we found that the MDET did not have a significant effect on the outcome of possessions greater than 15 seconds in length. This is similar to the trend we saw with RI and outcome; the effect appears to be transient. Thus in following analysis we consider only possessions less than 15 seconds. This includes possessions that end in free throws since a poor defensive effort can lead to defensive fouls. In total, we had 10,915 possessions. We computed the MDET score for each defensive possession and recorded a binary outcome (1 = opponent did not score, 0 = opponent scored or got to the free throw line). Figure 9 plots the MDET distribution for each outcome. Using the Kolmogorov-Smirnov test, we found that the two distributions were significantly different ($p < 0.01$). Looking at the distributions, we see that outcome 1's distribution is shifted toward lower MDET values than outcome 0. The mean MDET score for outcome 1 was 9.05 ft and the mean score for outcome 0 was 9.63 ft.

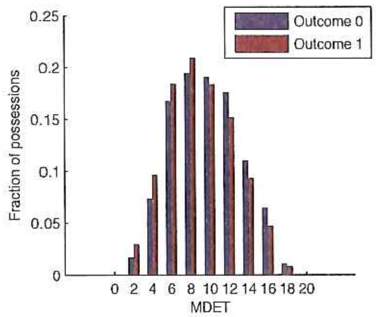


Figure 9: MDET distributions for outcomes 0 and 1.

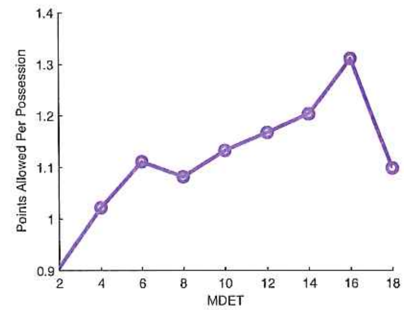


Figure 10: Points allowed per possession vs. MDET

These results are consistent with the intuition that better threat neutralization can prevent transition baskets on defense. Figure 10 displays the number of points allowed per possession versus MDET. An increase of MDET from 6 to 12, for example, leads to roughly 0.06 more points allowed per possession after a defensive rebound. Based on the dataset, there are about 25 possessions for a team per game that are continuous and follow a defensive rebound. It follows that a difference of 0.06 points per possession could change an opponent's score by 1.5 points

4 Early Threat Neutralization vs. Crashing/Retreating

Finally, we investigate how early threat neutralization relates to offensive rebound crashing. We hypothesized that having more players crash the boards will worsen threat neutralization performance since a player that crashes must run a longer distance to get back on defense. To test this hypothesis we considered all possessions in which a team missed a shot and did not get the offensive rebound and for which the following possession lasted at most 15 seconds. We then compare how the number of players crashing and retreating relates to MDET. Figure 11 shows that MDET tends to increase with an increase in RI and decrease with an increase in CI. These results are in agreement with our hypothesis that there is a tradeoff between offensive rebounding and getting back on defense.

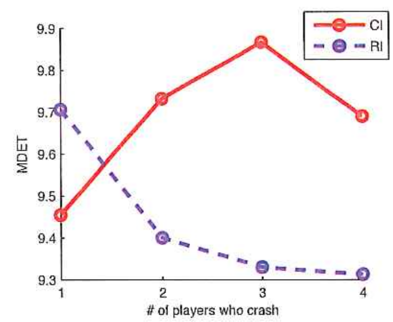


Figure 11: Comparison of CI and RI to MDET

5 Conclusion

In this paper we study optical tracking data in an attempt to quantify the tradeoff of going for the offensive rebound vs. getting ready for the transition to defense.

We analyzed the relationship between the movements of players when a shot is in the air and a team's ability to garner offensive rebounds. We defined two new metrics the Crash Index (CI) and Retreat Index (RI) that quantify the extent to which teams pursue an offensive rebound or ready themselves for transition defense. We also looked at what the offense does after the other team secures the defensive rebound. In doing so we introduced another metric, the maximum distance to early threats (MDET), designed to measure the effectiveness with which players defend during the transition period. We showed that there is a strong association between each of these three metrics and points/possession.

In conclusion, our results suggest that focusing on the offensive rebound immediately after the shot goes up seems to trump the gain a team gets with a head start on getting back. In the case of a defensive rebound by the other team early threat neutralization (as opposed to merely getting back early) can help limit the negative impact of transition baskets. The generalizability of these conclusions is limited by the data. For some teams we lacked data. Moreover, there are many factors we have yet to consider, e.g. the positioning of the defensive players, the game situation and especially the personnel on the floor.

Acknowledgements. We would like to thank STATS LLC for sharing their data, and Mike Zarren for all his help.

6 References

[1] Maheswaran, Rajiv, et al. "Deconstructing the Rebound with Optical Tracking Data". *MIT Sloan Sports Analytics Conference 2012.* March 2012.
[2] http://www.youtube.com/watch?v=PfhGqe3IoMU
[3] STATS SportsVu, http://www.sportvu.com/

Acceleration in the NBA: Towards an Algorithmic Taxonomy of Basketball Plays

Philip Maymin
NYU-Polytechnic Institute
Brooklyn, NY 11201
Email: philip@maymin.com

Abstract

I filter the 25-frames-per-second STATS/SportVu optical tracking data of 233 regular and post season 2011-2012 NBA games for half-court situations that begin when the last player crosses half-court and end when possession changes, resulting in a universe of more than 30,000 basketball plays, or about 130 per game. To categorize the plays algorithmically, I describe the requirements a suitable dynamic language must have to be both more concise and more precise than standard X's and O's chalk diagrams. The language specifies for each player their initial starting spots, trajectories, and timing, with iteration as needed. A key component is acceleration. To determine optimal starting spots, I compute burst locations on the court where players tend to accelerate or decelerate more than usual. Cluster analysis on those burst points compared to all points reveals a difference in which areas of the court see more intense action. The primary burst clusters appear to be the paint, the top of the key, and the extended elbow and wing area. I document the most frequently accelerating players, positions, and teams, as well as the likelihoods of acceleration and co-acceleration during a set play and other components intended to collectively lead to an algorithmic taxonomy.

1 Introduction

Basketball coaches preach and teach execution but objectively measuring execution, let alone estimating the contribution of execution on winning, has eluded analysis. Part of the problem is the language describing the desired execution. Basketball plays are routinely drawn up on chalkboards with standard static graphical notation, but the precise timing is often explained only orally to the huddled players. Here I introduce a dynamic algorithmic approach to concisely encode theoretical basketball plays and I describe its key characteristics. Important inputs to the language are the frequencies and locations of player acceleration.

I analyze optical tracking data on the 25-frames-per-second positional data of 233 regular season and post season 2011-2012 NBA games for half-court situations that begin when the last player crosses the half-court line and end when the offense no longer has possession. This subset, which by construction excludes both fast and secondary breaks, is ideal for analyzing set plays.

I document the teams, positions, and players in the dataset who exhibit the most and least frequent acceleration both on offense and on defense. Further, I evaluate the incidence of co-acceleration when multiple players experience bursts nearly simultaneously. I also determine the primary locations of bursts, compare and contrast them with the primary locations among inertial states, and evaluate the optimal number of such clusters. In addition, I provide new graphic tools, both static and dynamic, to ease analysis of these important issues. Taken together, these results build a roadmap towards an algorithmic taxonomy of basketball set plays.

Past research on optical tracking data in basketball includes [1], [2], and [3]. Acceleration in basketball has been studied from a physiological perspective, c.f. [4] in which the authors found that college basketball players were superior in terms of acceleration to non-athletes, but to my knowledge this is the first work exploring acceleration from optical data and its implications for an algorithmic approach to categorizing halfcourt set plays.

2 Data

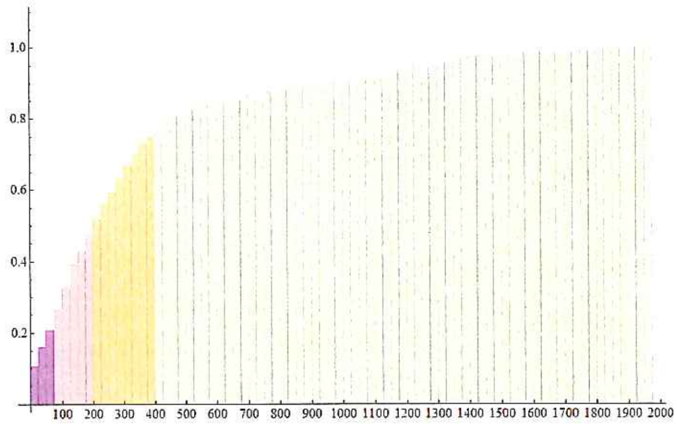
The three-dimensional SportVu optical tracking data from STATS LLC assigns to each player on the court an ordered (x, y) pair representing the position of their center of mass on a regulation 94 feet by 50 feet NBA court, and assigns to the ball an additional z coordinate specifying its height above the ground. These coordinates are recorded 25 times per second.

In addition, event identification information is automatically assigned to frames satisfying certain criteria such as dribbles, field goal attempts, and the like.

The data set covers 233 regular-season and post-season games during the lockout-shortened 2011-2012 season. Due to differential adoption, these games are skewed towards the teams that installed the required technology. Figure 1 shows the breakdown of home games for which the data was available.

Figure 1: Home Games in Sample, by Team

Figure 2: Cumulative Histogram of the Number of Plays per Player



I filter the sequences of these coordinates within games to create subsequences of halfcourt set plays, defined and implemented as follows. A halfcourt set play begins when all ten players are on the same side of the court, the opposite side from the previously recorded set play, and the ball is inbounds, within 20 feet of the halfcourt line, and nearest to one of the offensive players. The set play ends when the difference in time from the previous snapshot is more than 1/25 of a second (for example, a timeout has been called or the quarter has ended), when any player appears in the back court, or when the ball is nearest to a defensive player.

This definition is intended to capture choreographed set plays rather than improvised fast breaks or secondary breaks; in other words, possessions where each player's movements are the result of intentional practice. To focus on halfcourt set plays, out-of-bonds set plays are excluded, except for situations where the ball is thrown back inbounds to near halfcourt, at which point the definition above obtains and a presumptive regular set play can run.

The data set reduces to a universe of 30,950 plays lasting on average 180 frames each, or about 7 seconds. The data set includes location information for 10 players as well as the basketball, comprising more than 60 million coordinates in total. Some of the 456 distinct players are involved in more plays than others, because their team has more home games in the sample or because they have more playing time. Figure 2 shows the cumulative histogram of the number of plays each player is involved with in our universe. About three quarters of the players participate in at least one hundred distinct plays.

3 Methodology

Acceleration is computed as the second difference of the Euclidean distances between sequential moving average positions of a player on the court, divided by the standard gravity $g = 32.174\ ft/s^2$, multiplied by $625 = 25^2$ because the frames in the optical data are 1/25 of a second apart, and multiplied by 10 to express the acceleration in units of deci-g's, where 1 dg = 0.1 g is one-tenth of the standard acceleration of gravity g.

The window for the moving average, used to smooth the data for better accuracy in measuring the acceleration, is five. Finally, the most extreme accelerations are clipped because they are likely the result of random noise.

Specifically, the first smoothed position in a play for a particular player is calculated as the average of raw position numbers one through five for the player; the second position is the average of raw position numbers two through six; and so on. The acceleration is computed as the second difference, so the first acceleration is the second difference between the third and the first smoothed positions. Thus, it is the second difference between the average of raw positions three through seven and one through five.

In short, the computation for a single acceleration number requires eight frames from the optical data. While this may seem like a lot, it actually represents only 8 * 40 ms = 320 ms, which is literally in the blink of an eye: the experimentally measured blink duration is 334 ± 67 ms [5].

Bursts of high acceleration are rare. The table at the right of Figure 3 shows the conditional and unconditional frequency of occurrence as a function of the amount of acceleration. Acceleration is rare, and becomes rarer still for greater acceleration. Note that computed accelerations in excess of 7 dg, comprising less than one percent of the total, were clipped to 7 dg.

Figure 3 shows the frequency of acceleration on offense by team, with teams sorted in increasing order by their inertial proportion. The width of the bars represents the amount of data available for that team in our sample. The Houston Rockets spent the least time accelerating; the Boston Celtics the most. The corresponding team graph for defensive acceleration by team (not shown) has two substantive differences: the Milwaukee Bucks were even more inertial on defense than on offense, while the San Antonio Spurs moved up from the bottom into the middle of the pack.

Figure 3: Frequency of Offensive Acceleration by Team (Bar widths correspond to amount of data available in sample)

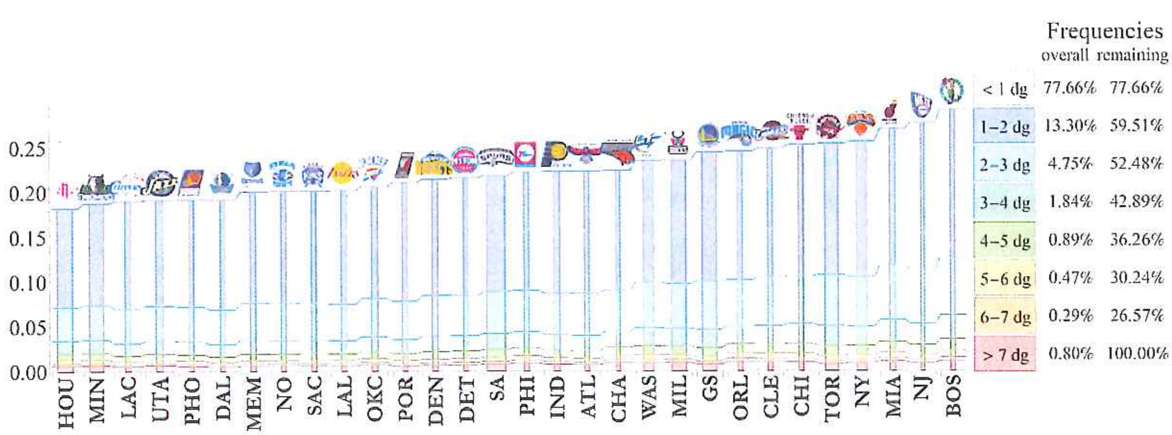


Table 1: Most and Least Frequent Accelerators per Position (Percent of Time > 1 dg)

Frequent Accelerators	Infrequent Accelerators
Solomon Alabi (41.98%) Tyson Chandler (35.05%)	Marcus Camby (18.46%) Marc Gasol (22.39%)
Joel Anthony (36.90%) Andrea Bargnani (34.90%)	Troy Murphy (17.81%) Tyrus Thomas (18.57%)
Kris Humphries (32.41%) Taj Gibson (31.12%)	Marcus Morris (12.05%) Matt Bonner (14.74%)
Carmelo Anthony (34.34%) LeBron James (32.27%)	Bobby Simmons (12.02%) Chandler Parsons (12.96%)
Paul Pierce (29.06%) Kevin Durant (26.20%)	Chase Budinger (11.02%) Danilo Gallinari (15.69%)
J.R. Smith (30.03%) Gary Forbes (27.94%)	Courtney Lee (10.43%) Danny Green (13.45%)
Kobe Bryant (26.82%) Ray Allen (25.56%)	Shannon Brown (13.64%) Daequan Cook (14.84%)
Toney Douglas (28.60%) Dwyane Wade (27.72%)	Gary Neal (14.38%) Daniel Gibson (14.55%)
Deron Williams (33.36%) Baron Davis (26.85%)	Derek Fisher (12.52%) Kyle Lowry (13.54%)

Figure 4: Frequency of Offensive Acceleration by Position

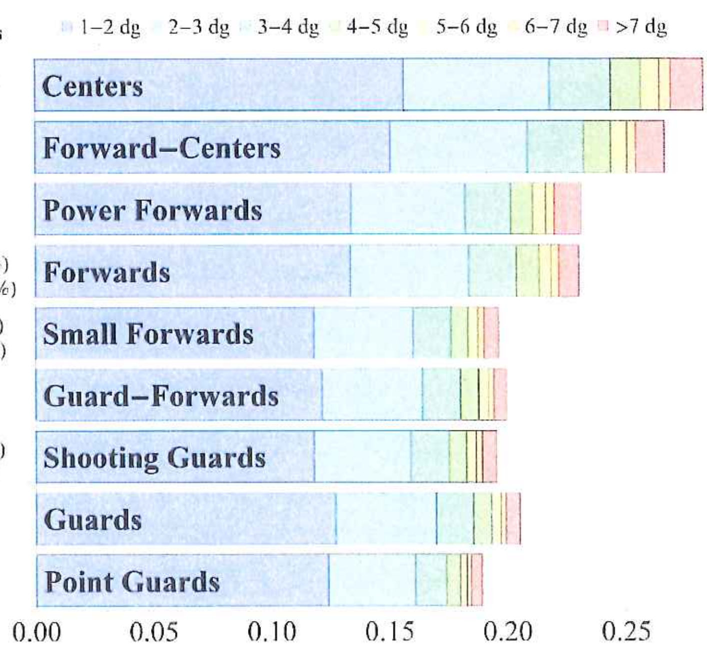


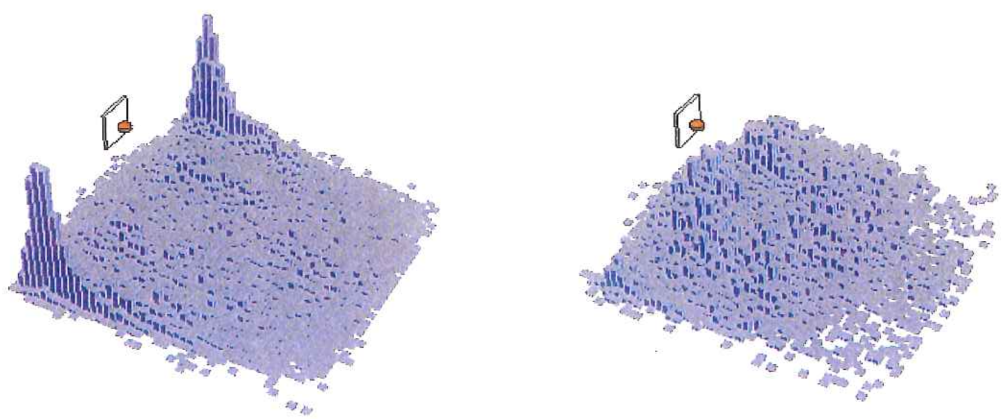
Figure 4 shows the frequency of acceleration on offense by position, with positions sorted by average height of players at that position. Centers have more bursts than guards, partially because they are more likely to set picks. Even among forwards and centers, bigger players tend to exhibit more extreme accelerations. Not shown are accelerations less than 1 dg in absolute magnitude; thus, point guards are not accelerating, i.e. inertial, 81 percent of the time while centers are inertial only 72 percent of the time. Note that inertial players are not necessarily standstill, they merely have constant velocity. The corresponding graph for defensive acceleration by position (not shown) does not substantially differ.

Table 1 lists the players with the highest and the lowest frequency of accelerating at least 1 dg while on offense (in other words, those with the lowest frequency are the ones with the highest inertia). The table would look somewhat different for different thresholds of acceleration, e.g. if restricting only to accelerations greater than 3 dg instead of 1 dg, but the top 10 names in each category tend to be relatively stable. Further, essentially the same names appear on the corresponding defensive table (not shown): players seem to accelerate, or not, based on who they are, not based on whom they are guarding.

The spatial distribution of all players differs markedly from accelerating players. Figure 5 displays the three-dimensional histogram of halfcourt player locations for a random subsample of all players, and for a random subsample of players at times when they are accelerating or decelerating at a magnitude of 5 dg or greater. Random subsamples were used to facilitate display: the entire universe of halfcourt set plays contains nearly 30 million offensive player positions.

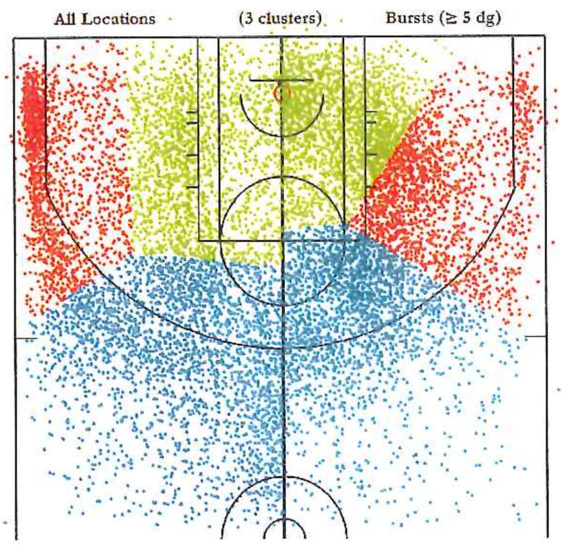
Figure 3: Histogram of Positions of All Players and Accelerating Players

All Locations (Acceleration ≥ 0 dg) | **Burst Locations (Acceleration ≥ 5 dg)**



A cluster analysis further highlights the differences while also suggesting common burst areas. Figure 6 shows these results. Note the differences between the left-hand sides and right-hand sides of each halfcourt. As with Figure 5, the corner three is a popular location, albeit not one with much acceleration. Among burst points, the three primary clusters appear to be the paint, the top of the key, and to a lesser extent, the combined area of the extended elbows and wings. Extending to five clusters separates out the elbow and the corner wings as additional areas, but do not appear to be as well demarcated as the three clusters. Thus, we are justified in treating the three cluster graph as an appropriate model for burst points, which will become starting points in our language.

Figure 4: Cluster Analysis of Positions of All Players and Accelerating Players



4 Play Language and Specification

Bursts of extreme acceleration tend to happen in the paint, at the top of the key, or in the extended elbow and wing areas. With each area on the left and on the right of the court, there are six possible starting and ending spots for player trajectories:

1) **LP** and **RP**: Left Paint and Right Paint
2) **LK** and **RK**: Left Key and Right Key
3) **LW** and **RW**: Left Wing and Right Wing

In principle, player trajectories may happen simultaneously or after previous trajectories are finished. Based on the standard "X's and O's" graphical notation that does not specify an order for trajectories, it indeed appears as if all movements happen simultaneously. In practice, we can explore how often burst points happen at the same time.

How often do multiple players accelerate simultaneously? If by "simultaneously" we mean the exact same frame (1/25 of a second), then the answer is virtually never. But if by "simultaneously" we mean "within one second of each other," then we can count the number of times within all rolling 25-frame subperiods that no players, exactly 1 player, exactly 2 players, exactly 3 players, exactly 4 players, or all 5 players were accelerating.

Figure 5: Proportions of Co-Accelerating Players

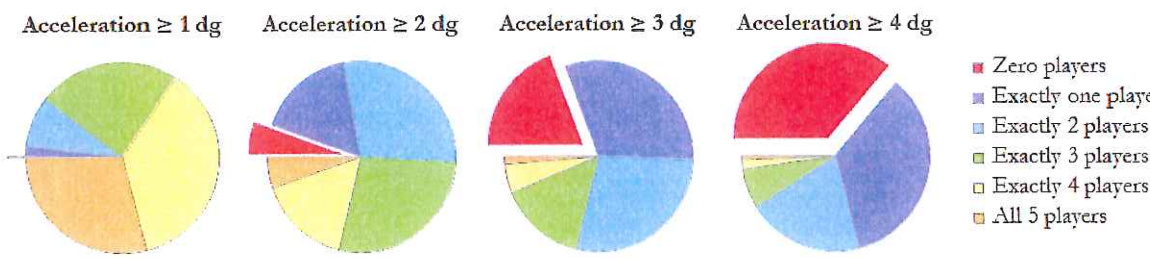


Figure 7 shows the pie charts of these counts for acceleration thresholds of 1 dg, 2 dg, 3 dg, and 4 dg. The likelihood of multiple players accelerating or decelerating at a magnitude of at least 1 dg in any given one-second interval is nearly 100 percent: in other words, mild acceleration is the norm for most players most of the time. The story changes at more extreme bursts, however. The co-acceleration likelihood drops to about 75 percent for 2 dg, 50 percent for 3 dg, and 25 percent for 4 dg.

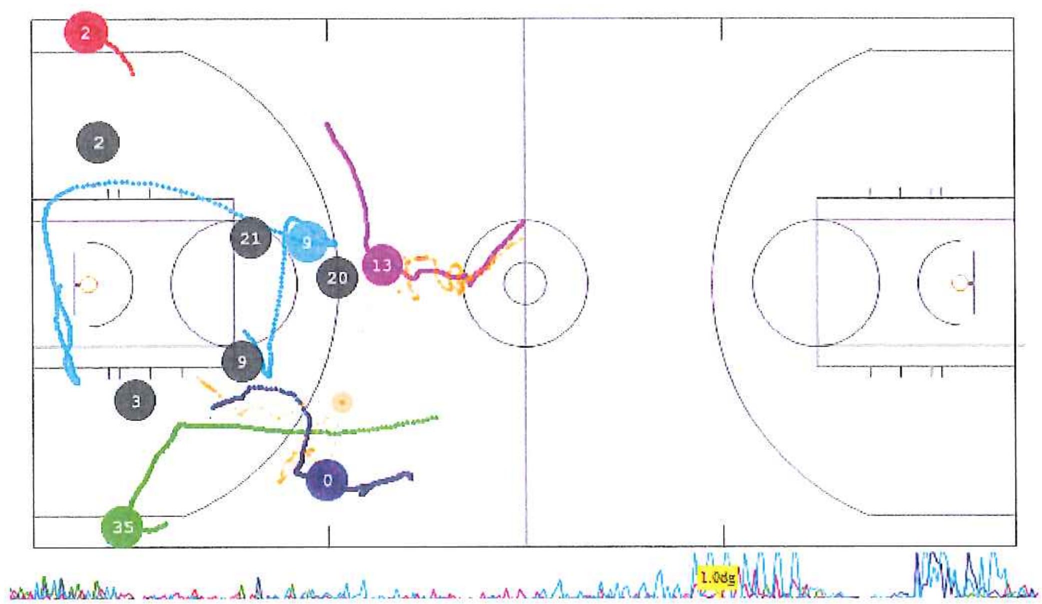
It may be useful and simplifying to assume that no player begins a new trajectory while another player is in the middle of his trajectory. This is a relatively non-constricting assumption because trajectories may be broken up into parts, e.g. a player crossing the court from one wing to the other may, if necessary, be modeled as running first to the nearest painted area, then to the other, and then finally to the opposite wing.

Therefore, the state of a play at any given time requires noting for each player which burst cluster they are currently in and which one they are aiming towards. If they are not accelerating, then they may be said to aiming to the same cluster they are currently in. Finally, the basketball itself needs to be modeled as well; because we also need to know only its current location and destination, it can be treated as a sixth player on the court.

An example may help illustrate the approach. Figure 8 shows the snapshot from a video examination of the final assisted field goal of the Thunder against the Spurs in Game 6 of the 2012 Western Conference Finals. The play started with a little under two minutes remaining as James Harden (#13) dribbled over halfcourt, defended by Manu Ginobili (#20). The snapshot in Figure 8 occurs about two-thirds of the way through the play, at a time after Harden has passed the ball to Russell Westbrook (#0) but before Westbrook has caught it.

Note the five individual offensive trajectories. Thabo Sefolosha (#2) exhibits essentially no acceleration or velocity in RW. Kevin Durant (#35) accelerates from LK to LP then to LW. Harden accelerates from RK to RW though he doesn't quite reach it before the play ends. (However his acceleration is nevertheless important as it helps draw Ginobili away from the action.) Serge Ibaka (#9) has the most complicated route and the most amount of screens, accelerating from LP to RP, then to RK, then finally to LK, where his final pick frees up Westbrook to take and make the jumper and extend the Thunder's lead to four points, essentially sealing their victory in the game and the series. Note also that Westbrook's acceleration towards the end of the play coincides with Ibaka's; similarly, Harden's acceleration starting at the snapshot shown also coincides with Ibaka's.

Figure 6: 2012 Western Conference Finals, Game 6, OKC@SA, Q4 1:49 – 1:36



5 Conclusions

Calculating and analyzing the accelerations of each offensive and defensive player in each halfcourt set from optical tracking data, along with novel static and dynamic visualizations, helps shed light on the rare but critical bursts that help define basketball plays and suggests a road towards an algorithmic description and ultimately a taxonomy of plays. Future research directions may include attempting to predict future accelerations from past data alone, implementation of a tool to help advance scouts categorize plays, and exploring the relationships between acceleration and co-acceleration on field goal percentage, spacing, and execution variability.

Acknowledgments

I am grateful to Ken Catanella, David Griffin, Sam Hinkie, Brian Kopp, Allan Maymin, Daryl Morey, Eugene Shen, and Ryan Warkins for discussions around the concepts addressed in this paper, and to STATS LLC for their data.

References

[1] Sandy Weil, "The Importance of Being Open: What Optical Tracking Data Can Say About NBA Field Goal Shooting," MIT Sloan Sports Analytics Conference, presentation at http://sloansportsconference.com/?p=661, Mar. 2011.

[2] Allan Maymin, et al., "Individual Factors of Successful Free Throw Shooting," Journal of Quantitative Analysis in Sports, vol. 8, no. 3, Oct. 2012.

[3] Rajiv Maheswaran, et al., "Deconstructing the Rebound with Optical Tracking Data," MIT Sloan Sports Analytics Conference Proceedings, Mar. 2012.

[4] Gillam G. McKenzie, "Identification of anthropometric and physiological characteristics relative to participation in college basketball," National Strength & Conditioning Association Journal, vol. 7, no. 3, pp. 34-36, Jun. 1985.

[5] Frans VanderWerf, et al., "Eyelid Movements: Behavioral Studies of Blinking in Humans Under Different Stimulus Conditions," Journal of Neurophysiology, vol. 89, no. 5, pp. 2784-2796, May 2003.

7th ANNUAL
MIT SLOAN
SPORTS ANALYTICS CONFERENCE
MARCH 1-2, 2013 BOSTON CONVENTION AND EXHIBITION CENTER
presented by ESPN

Live by the Three, Die by the Three? The Price of Risk in the NBA

Matthew Goldman
UC San Diego
Dept. of Economics
La Jolla, CA 92093
mrgoldman@ucsd.edu

Justin M. Rao
Microsoft Research
102 Madison Ave.
New York, NY 10016
justin.rao@microsoft.com

Abstract

An important problem facing a basketball team is determining the right proportion of 2 and 3 point shots to take. With many possessions remaining, a team should maximize points—a 3-pointer is simply worth 1.5 2-pointers. 3-point attempts have roughly double the per-shot variance as 2-point attempts, but a team should be "risk neutral." As time remaining decreases, the trailing team *should* place an increasingly positive value on risk; the opposite holds for the leading team. Our game theoretic analysis yields a testable optimality condition: 3-point success rate *must* fall relative to 2-point success rate when a team's preference for risk increases. Using four years of play-by-play data, we find strong evidence this condition holds for the trailing team only. As a lead decreases, the leading team should become more risk-neutral, but teams in this circumstance actually tighten up and become more risk averse, contrary to what their risk preferences ought to be to maximize the chance of winning the game. We also show that if the offense shoots more 3's as it becomes risk-loving this implies the attack can be varied more readily than the defensive adjustment. 3-point usage does increase with the trail team's preference for risk, but actually falls for the leading team. Teams get it right when losing and wrong when winning. We also find a strong motivating effect of losing—the trailing teams displays an overall boost in efficiency for both shot types.

2013 Research Paper Competion
Presented by:
42ANALYTICS
SAP

1 Introduction

In order to optimize, a basketball team must determine the right proportion of 2 and 3 point shots to take. In this paper we study how NBA teams solve this problem. The trade-off between 2's and 3's depends critically on the score margin and the time-remaining in the game. Early in the game, a team should maximize points per possession—a 3 pointer is simply worth 1.5 2-pointers. Since 3-pointers have about double the per-shot variance in point outcomes as 2-pointers, this implies a team should be "risk neutral" in these situations. Moving towards the end of the game changes this trade-off. With a relatively small number future opportunities, the trailing team should place an increasingly positive value on risky 3-point opportunities because they need a large swing in points to catch up; conversely the leading team should place a negative value on risk, favoring predictable scoring opportunities. Using detailed play-by-play data for six years of NBA games, we empirically quantify the true value of 3-pointers relative to 2-pointers as a function of score margin and time remaining. We do so by calculating the impact a made shot of each type has on the chance the team wins the game for all the game states in our sample. It is not uncommon for a made 3-pointer to be worth as much as 1.8 and as little as 1.2 times the "win value" of a made 2-pointer.

We model a team's choice of the proportion of shots between 2 and 3-pointers using the tools of game theory. Solving the model gives the optimal mix of 2's and 3's. If we assume the defense cannot adjust, then when the offense's preference for risk increases, optimality implies it shoots more 3-pointers, 3-pointer efficiency falls, 2-pointer efficiency rises and the win value of 3's rises. The model makes clear that even in the risk-neutral setting, optimal shot selection does not imply that 2 and 3 pointers offer the same average points per attempt (average efficiency).

We extend the model by allowing for the defense to adjust the allocation of scarce defensive attention between two and three-point defense. An increasing preference for risky 3-pointers by the offense is associated with an increase in the defense's incentive to guard against them. Allowing for a flexible class of defensive responses, we show that 3-point efficiency should be inversely correlated with a team's preference for risk—as a team becomes more risk loving, optimal shot selection implies that 3-point percentage *must* go down. Our final extension allows for a direct motivating effect of trailing; in this case the key prediction is the 3-point percentage must fall relative to 2-point percentage. The defensive adjustment model shows that the offense shoots more 3's when their preference for risk goes up only if the offense's ability to vary the"attack" exceeds the defense's ability to adjust.

We empirically test these predictions using four seasons of NBA play-by-play data. We allow for different model estimates for each team-season and condition non-parametrically on both the offensive and defensive 5-man lineup to control for potential biases induced by substitution. Consistent with optimal shot selection, the trailing team exhibits strong statistical evidence ($t = 8.28$) that the point value of 3's falls relative to 2's as the offense's preference for risk increases. The trailing team also shoots more 3's. In stark contrast, the leading team significantly violates our key optimality proposition. Leading teams shoot *fewer* 3's as their preference for risk increases and these 3's have *higher* point value. As a lead decreases, the leading team should become more risk-neutral, but instead tighten up and actually become more risk averse, contrary to what their risk preferences ought to be to maximize the chance of winning the game.

Interestingly, we also find a large motivating effect of being behind (also discussed in our related paper [4])—for a given offensive and defensive line-up, the trailing team displays an increase in efficiency for both 2 and 3-pointers. This effect is similar in spirit to the findings of Pope and Berger (2011) [1] that a team trailing by 1-point at halftime wins slightly more often than the team leading by 1-point. The extra motivation of being behind has ties to Kahneman and Tversky (1979)'s theory of "loss aversion"—the principle that people are more motivated to "remove losses" than "seek gains" [8].[1] In the context of the NBA, our results indicate that these sort of motivational links do not depend on a halftime speech, tactical adjustments or line-up changes. Whether this effect is driven by the leading team's complacency or the trailing team's motivation (or both) is not a question our data can speak to, but the net effect is clear.

The loss-aversion finding combined with the suboptimal shot selection of leading teams helps explain why teams tend to stage more comebacks than we'd otherwise expect. Since games tend to get close late, clutch moments are

[1] Rick and Loewenstein (2008) [6] also provide laboratory evidence in favor of this type of "motivational loss aversion."

relatively frequent. We show that for an average team it's harder to score in clutch moments, but very good offenses actually do better in the clutch and bad defenses get worse. Taken together, this means good teams have an even greater advantage when the chips are down.

2 Quantifying a Team's Objective Function

A team's goal is to win the game. Accordingly, we are interested in estimating the mathematical function that gives the "win value" of a given action. "Win value" refers to the impact the action has on the probability a team wins the game. The three most important factors that determine the win value of an action at a given game state, especially late in the game, are the score margin, time remaining and possession of the ball. The increase in win probability of adding 2 or 3 points to the team's current score are denoted WV_2 and WV_3 respectively. The most straightforward estimation approach for these quantities is to take a large number of games at a given game state X and compare the probability of winning at X to a "nearby" state X'. For example, natural variation in actions taken at X give us some cases where a shot was missed, some where a 2-pointer was made and some where a 3-pointer was made. Intuitively, comparing the frequency which a team wins after these outcomes gives us the value these actions.

We define $p(X)$ as the probability a team wins the game in state X. Econometrically we have two options to calculate this quantity. The first method is "non-parametric"—it relies on local averages as described in the above paragraph. The second is the "parametric" procedure developed in Goldman and Rao (2012) [3]. We describe it here in the Appendix for completeness. This approach conditions on team quality, home court and game-state using a Probit regression. Appendix Figure 1 shows that both methods yield very similar projections. Given the lower noise in the parametric estimates (smoother function, Panel 1), our analysis will proceed using those estimates.

The relationship between the win value of 3's and 2's can be represented by the parameter α defined as:

$$\alpha = \frac{WV_3}{WV_2}. \tag{1}$$

α defines the degree to which 3-pointer win value diverges from 1.5 2-pointers. We refer to standard point values as "nominal values." In a most game situations, especially in the first half, scoring 3 points on a given possession is worth very close to 1.5 times scoring 2 points. When $\alpha > 1.5$, the win value of a 3-pointer exceeds it's nominal value. This occurs for the trailing team, especially late in the game. The effect can be seen in the convexity in the trailing region of Appendix Figure 1. The reason is that the trailing team needs a relatively large swing in points to catch-up—a higher variability shot is worth more because it increases the chance of this large swing. The opposite is true for the leading team (which has $\alpha < 1.5$)—here a 3-pointer is worth relatively less than usual since the team should be risk-averse.

In Figure 1 Panel 1 we plot α as function of game state (margin, time remaining) for even strength teams on a neutral court over the first 3 quarters (Panel 1) and the fourth quarter (Panel 2, note the change in y-axis scale). In the first half α is always close to 1.5. In the third quarter we see more variation; α is between 1.4 and 1.6 provided the margin is less than 11 points. In the 4th quarter, α varies widely. With fewer possessions remaining, the trailing (leading) team's preference for risk increases (decreases) dramatically. [2]

3 Modeling a Team's Shot Allocation Problem

We express a team's optimization problem as a function of α, solving gives the optimal mix of 2 and 3-pointers for each game situation. We will employ a concept in basketball analysis called the "usage curve" [5, 7, 2]. The

[2] α is a natural proxy for a team's preference for risk because it maps directly to the relative preferences over potential outcomes. Consider a case where $p_2 = 0.50$ and $p_3 = 0.33$. In this case there is an expected nominal value of one point for reach shot. The variance in the return of a 3-point attempt is $3^2.33 * .66 = 1.96$ and for the two-pointer $2^2.5^2 = 1$. Suppose $\alpha = 1.7$. This means the expected utility (expected real value) of a 3-pointer is $0.33*1.7*WV_2=0.561*WV_2$ and 2-pointer is worth $0.5 * WV_2$; or in other words, the 3-pointer is worth 12% more in win value, despite having equal nominal value. If we model the team as have preference over mean and variance, we could map any α to a utility value of variance. However, we view α as a more directly interpretable parameter.

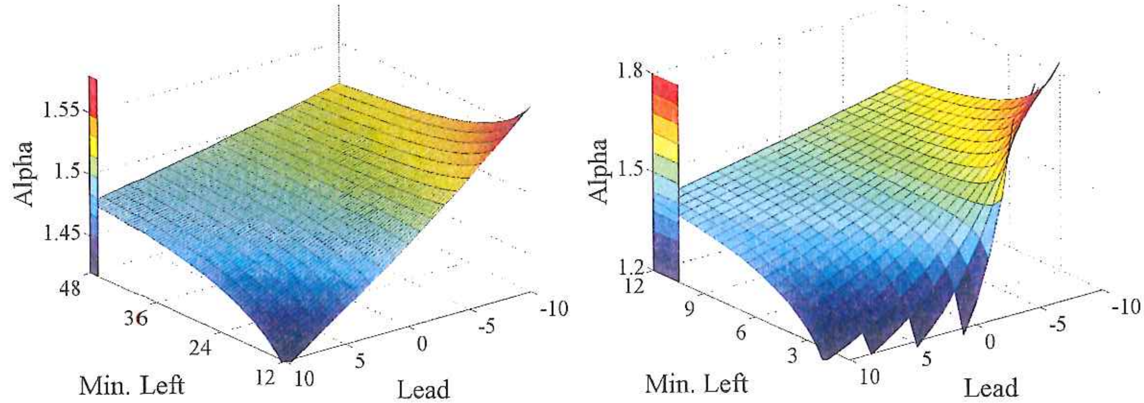


Figure 1: α as a function of game state. Quarters 1-3 (left) and Quarter 4 (right).

usage curve relates the frequency of a given shot (in this case a 2 or 3-pointer) to it's success rate. Usage curves are naturally assumed to be downward sloping and were estimated as such in Goldman and Rao (2011). This feature implies that as a team shoots more 3's, the probability of success on each successive 3-point attempt goes down.[3] Let $\phi(u_3)$ denote the average probability of success for 3-pointers when the fraction of shots attempted as 3's is $u_3 \in [0, 1]$ and $\psi(1 - u_3)$ the corresponding average probability of success on 2-poin attempts. The downward sloping assumption implies $\frac{d\phi}{du_3} < 0, \frac{d\psi}{du_3} > 0$—as a team shoots more 3-pointers the average returns to 3's go down and the average returns to 2's goes up. Recalling that we define the win value of 3's in relation to 2's as $WV_3 = \alpha WV_2$, we can now write the team's maximization problem as:

$$\max_{u_3} u_3 * \phi(u_3) * \alpha WV_2 + (1 - u_3) * \psi(1 - u_3) * WV_2 \qquad (2)$$

the first order condition can be rearranged to give:

$$\alpha(\phi(u_3) + \phi'(u_3) * u_3) = \beta(1 - u_3) + \psi'(1 - u_3)(1 - u_3) \qquad (3)$$

The first order condition states that the marginal returns to 2-pointers and 3-pointers should be equal. The left side gives the marginal returns to shooting a 3-pointer. Shooting an extra 3-pointer returns the current average value $\alpha * \phi(u_3)$, but it also impacts the average value of all the other 3-pointers taken u_3 by a degree given by the slope of the usage curve ($\phi'(u_3)$). The right hand side gives the marginal returns to shooting a 2-pointer and can be understood with similar logic. In this model with no defensive adjustment, u_3 is increasing with α. To see this note that if α increases then the left side goes up because the term $\phi(u_3) + \phi'(u_3) * u_3$ has to be positive, otherwise the marginal 3-pointer nets negative value. So the left side must increase, to counter-act this, the right side must go up as well which occurs only when u_3 increases.

Appendix Figure 2 gives a graphical representation of the maximization problem and the impact of an increase in α. Starting at $\alpha = 1.5$, point A determines the baseline optimal shot mix. Optimal shot choice does not imply 2-pointers and 3-pointers offer the same average point value. The difference in average shot value is determined by the slope of the usage curves and the y-intercepts. In practice 3-pointers tend to return greater average efficiency and are shot less often than 2's, together this imples a higher y-intercept and a steeper slope in the usage curve for 3's. The figure also shows the impact of an increase in α, which is represented by the shift out in the 3-point value

[3] Goldman and Rao (2011) give an empirical reason for downward sloping usage curves in this setting. If we model the offense as getting shot opportunity arrivals over the course of a 24-second shot-clock, then to take more 3's, the team has to accept lower quality 3-point opportunities on the margin.

curve. The new equilibrium is given by the vertical line intersecting A'. Point B' gives the new win value of 2's and point C' the new win value of 3's. Point D gives the new nominal value of 3's (the point value). We are now now in a position to state our first proposition:

Proposition 1 *In the model with no defensive adjustment, as α increases the fraction of 3's attempted (u_3) goes up, the nominal value of attempted 3's goes down, the nominal value of attempted 2's goes up and the real value of attempted 3's goes up.*

Proof: See Appendix

The model without defensive adjustment is useful to provide intuition and also can be interpreted as representing a world in which defensive adjustments matter relatively little, which may apply, for instance, to a team playing man-to-man defense that lacks the quickness and length to alter their strategy and really clamp down on opposing 3-point shooters. Incorporating defensive adjustments to our model is straightforward; the defense's objective is simply the opposite of the offense's (it wants to minimize equation 2)—an increase in the value of 3's increases the incentive to defend against them. We assume the defense has a unit of "defensive resources," which it can apply to defending 2's and 3's: $d_2 + d_3 = 1$; more defensive attention lowers the success rate of a shot type. We modify the usage curves to include defense ($\phi(u_3, d_3), \beta(u_2, d_2)$). Analysis of this model is involved so we have placed it to the Appendix. Interested readers are directed there. We now state our second and third propositions:

Proposition 2 *In the model with defensive adjustment, as α increases the nominal value of attempted 3's falls and the nominal value of attempted 2's rises.*

Proof: See Appendix

Proposition 3 *In the model with defensive adjustment, as α increases the change in the usage rate of 3's is ambiguous. It depends on the slope of the 3-point usage curve, the impact of defensive on the marginal shot values and the concavity of the usage curves with respect to defensive pressure.*

Proof: See Appendix

Proposition 2 states that the prediction of the no-defense model that carries through is the drop in the nominal efficiency of 3's as α increases. Proposition 3 states that the other predictions of the simple model are not robust when we allow for a large class of defensive pressure adjustments. With defensive adjustment, the offense will shoot more 3's provided the defense cannot adjust pressure effectively enough to discourage these additional attempts. The details, which are in the Appendix, are a bit a hard to grasp at first, but the main intuition comes down to relative flexibilities of the offensive attack and defensive response.

Our final extension of the model is to allow for a multiplicative function of α on each usage curve that accounts for a possible motivational impact of being behind in the game (we could also model this as an additively separable term). It is easy to show that this term will cancel out of the first order conditions. However, we must amend Proposition 2 to be:

Proposition 4 *If we allow for a motivational impact of trailing and defensive adjustment, then as α increases the nominal value of 3's falls **relative** to the nominal value of 2's.*

When 3's become more valuable, maximization implies that the efficiency of 3's must fall relative to 2's. This is our most robust prediction, as it is true in the very general defensive adjustment setting and allowing for an extra motivational impact of being behind. Optimization *requires* that Proposition 4 holds.

4 Results

In our empirical analysis, we are careful to exclude situations in which one team has less than a 5% chance of winning. Actions in "garbage time" lack meaningful consequences and tell us nothing useful about whether a team is optimizing. We also eliminate fast-breaks (shot clock > 14) and end of quarter shots, as they tend to have very deifferent strategic considerations.

4.1 Frequency of 3-point vs. 2-point shot attempts

We first examine the impact of α on the usage rate of 3-pointers. We model the probability a team's first shot on a possession is a 3-pointer using a random coefficient linear probability model, which allows coefficients to vary for each team in each season (a "team-year"). We control for unique five-man offensive lineup (δ) and opposing defensive lineup (γ) using fixed effects for each line-up. Our estimating equation is given by:

$$Pr(3\text{PA}_p) = \delta_{Off_p} + \gamma_{Def_p} + \beta_{1,t}[\alpha_p \times 1\{\alpha_p \leq 1.5\}] + \beta_{2,t}[\alpha_p \times 1\{\alpha_p > 1.5\}],$$

where Off_p and Def_p denotes the five-man offensive and defensive line-ups on possession p, respectively, and α_p denotes the value of α faced by the offensive team on possession p. This specification is very general and ensures that we are not confounded by lineup effects. β_1 gives the impact of an increase in α for possessions when $\alpha < 1.5$, which corresponds to the case when the team is leading. In this case, α increasing pushes the team closer to the risk-neutral baseline. β_2 gives the impact of an increase of α when $\alpha > 1.5$, meaning the team is trailing. In this case, α increasing moves the team to a more desperate, risk-loving situation. In both cases, as α increases, the team's preference for 3-pointers is increases relative to 2-pointers. Estimating this model for each team-year in our sample produces 120 total estimates of each parameter.

Table 1: Random−coefficient estimates of the impact of α on three−point usage rates.

Explanatory Variable	Weighted average† coefficient *t−stat*	Mean coefficient *t−stat*	Median coefficient *t−stat*‡
$\beta_1 : \alpha_p \times 1\{\alpha_p \leq 1.5\}$	−0.105	−0.108	−0.0799
	t=−8.21	*t=−8.24*	*t=−2.37*
$\beta_2 : \alpha_p \times 1\{\alpha_p > 1.5\}$	0.231	0.243	0.175
	t=13.40	*t=13.44*	*t=3.47*

Team-years= 120, Shots=481,544
† Inverse variance weights used to aggregate coefficients.
‡ Sign test used to construct t−statistics on the median.

Table 1 aggregates the estimated coefficients. Examining the first row, we see that β_1 is significantly *negative*. When a leading team's α increases, it shoots fewer 3-pointers; that is, the opposite direction as predicted by the no-defensive adjustment model. As shown in Figure 1, α increasing for the leading team means, all else equal, the game is getting closer. The leading team appears to tighten up in this situation, shooting fewer 3's, despite the fact their preference for 3's is increasing. Taken alone, we cannot conclude that this pattern of behavior, while perhaps surprising, is a violation of optimal shot selection because in our model with defensive adjustments that the impact of α on 3-point usage rate depends on the relative adjustment ability of the offense vs. defense. However the estimates of β_2 give us some evidence that this negative coefficient is in fact a sign of sup-optimal behavior. Examining the second row, we see that for the trailing team, as α increases the team shoots more 3's.

The usage behavior displays an interesting asymmetry. When a trailing team's preference for 3's goes up (becomes more risk-loving), it shoots more 3's and fewer 2's. But when a leading team should become more risk-neutral, it actually behaves in a more risk-averse way, switching to 2-pointers. If we make the reasonable assumption that the defensive adjustment technology is similar when a team is leading vs. trailing, then the usage estimates indicate that trailing teams respect changing risk preferences (the "price of risk"), but leading teams do not, even so far as inverting the relationship.

4.2 The efficiency of 3-point vs. 2-point shot attempts

We delve further into this asymmetry in our analysis of shooting efficiency. Recall that our most robust prediction is given by Proposition 4. Even if players get generally better when they are trailing, our theoretical model still implies that 3-point opportunities cannot increase in value as much as 2-point opportunities. That is, the gap in point value between 3 and 2-point attempts must be declining with α. We investigate this claim with the following random-coefficients linear regression model:

$$E(\text{Points}_p) = \delta_{Off_p} + \gamma_{Def_p} + \beta_{1,t} \cdot 1\{3\text{PA}_p\} + \beta_{2,t}[(\alpha_p - 1.5) \times 1\{\alpha_p \leq 1.5\}] + \beta_{3,t}[(\alpha_p - 1.5) \times 1\{\alpha_p > 1.5\}]$$
$$+ \beta_{4,t}[1\{3\text{PA}_p\} \times (\alpha_p - 1.5) \times 1\{\alpha_p \leq 1.5\}] + \beta_{5,t}[1\{3\text{PA}_p\} \times (\alpha_p - 1.5) \times 1\{\alpha_p > 1.5\}].$$

We again include fixed effects for each unique five-man offensive and defensive line-up to exclude confounding effects from lineup selection. We have written expected points as the dependent variable, but we actually use 3 different, albeit similar, measures (we will use the word "efficiency" to refer to the class of dependent variables we use and discuss differences below). β_1 can be interpreted as the average efficiency differential between 3 and 2-point shots in a risk-neutral ($\alpha = 1.5$) game state. β_2 captures the impact of α on 2-pointer efficiency for the leading team ($\alpha \leq 1.5$), while β_3 captures this effect for the losing team. β_4 and β_5 directly test Proposition 4, these coefficients represent the *differential* effect of α on the efficiency of 3-point attempts for a winning and losing team respectively.

The three dependent variables are reported in Panels 1–3 of Table 2. Panel 1 gives the"effective field goal %," which is simply the points scored on the shot for shooters that were not fouled. Panel 2 gives "true shooting %," which takes the number of points scored on the shot plus any free-throws made related to the shot[4]. In Panel 3 we report "gross offensive efficiency," which is the number of points scored on a possession after the shot attempt occurs (this includes the shot going in, free-throws related to the shot and any points scored after an offensive rebound(s)). By using all three variables we can discern if the effects are being driven by differential offensive rebounding rates across shot types or fouling patterns by the defense.

Estimates of β_{1-5} are computed for each team-year. The following results apply to all three dependent variables, we discuss differences where necessary. β_1 is strongly positive—3 pointers have higher average point returns in the risk-neutral baseline. For gross-offensive efficiency and true shooting %, the mean estimate is about 0.15 points-per-shot. Recall that this implies a higher constant and steeper slope for the 3-point usage curve. The estimates for β_2 and β_3 are dramatically, and very significantly, positive. This means that as a team goes from being ahead to being behind, 2-pointers get more and more efficient. This effect is in fact stronger for a trailing team ($\beta_3 > \beta_2$). This is strong evidence of the motivational impact of losing.[5] We note that the $\beta_3 : \beta_2$ ratio is highest for gross possession efficiency, indicating that some of the motivational effect of losing is coming through offensive rebounds (which the two other measures ignore, the estimates indicate about 30% of the effect is coming through rebounding).

The key test of optimality lies in the estimates of β_4 and β_5. Proposition 4 states that optimal response to changing incentives over risk requires that both coefficients are negative. The reason is that as α increases, a offense should become more risk-loving and the defense should want to defend 3's more—since the true value of a 3-pointer has increased, the nominal or point value of a 3 should decrease in equilibrium. First the positive results: this optimality condition is met for trailing teams, β_5 is significantly negative ($p < 0.0001$ for the weighted average in all specifications). As the trailing team becomes increasingly risk-loving, the point value of 3-pointers declines relative to 2-pointers just as predicted by the theory. Recall from Table 1 that the trailing team also responds by

[4]Note these first two measures are usually computed with a 1 for a made 2-pointer and 1.5 for a made 3-pointer in order to have the same scale as traditional FG%. We instead use 2 and 3 respectively (effectively doubling the value), in order to stay consistent with our other third measure.

[5]The impact of losing on player motivation and performance is examined more closely in a related paper of ours [4].

Table 2: Random–coefficient estimates of the impact of α_p on nominal returns to three point attempts.

	Effective Field Goal %			True Shooting %			Gross Possession Efficiency		
Explanatory Variable	Weighted† avg. coeff.	Mean coeff.	Med.‡ coeff.	Weighted† avg. coeff.	Mean coeff.	Med.‡ coeff.	Weighted† avg. coeff.	Mean coeff.	Med.‡ coeff.
$\beta_1 : 1\{3PA_p\}$	0.223 t=49.50	0.221 t=48.59	0.218 t=10.77	0.143 t=31.95	0.142 t=31.40	0.143 t=10.22	0.155 t=34.74	0.155 t=34.20	0.158 t=10.59
$\beta_2 : \alpha_p^* \times 1\{\alpha_p \leq 1.5\}$	1.71 t=39.72	1.78 t=40.21	1.65 t=10.22	1.58 t=39.04	1.65 t=39.76	1.54 t=10.22	1.84 t=45.54	1.92 t=46.14	1.76 t=10.22
$\beta_3 : \alpha_p \times 1\{\alpha_p > 1.5\}$	2.3 t=38.20	2.34 t=37.05	2.14 t=10.22	2.8 t=50.82	2.83 t=49.16	2.64 t=10.95	3.33 t=60.30	3.35 t=58.26	3.25 t=10.77
$\beta_4 : 1\{3PA_p\} \times \alpha_p \times 1\{\alpha_p \leq 1.5\}$	0.213 t=2.44	0.185 t=2.03	0.142 t=0.91	0.343 t=3.97	0.337 t=3.74	0.348 t=2.56	0.387 t=4.47	0.384 t=4.26	0.317 t=2.37
$\beta_5 : 1\{3PA_p\} \times \alpha_p \times 1\{\alpha_p > 1.5\}$	−0.71 t=−5.92	−0.758 t=−6.11	−0.66 t=−2.92	−1.1 t=−9.45	−1.18 t=−9.73	−0.965 t=−4.56	−0.969 t=−8.28	−1.03 t=−8.47	−1.05 t=−4.56

Team-years = 120, Shots = 481,544, † Inverse variance weights used to aggregate coefficients
‡ Sign test used to construct t−statistics on the median. * We suppress the −1.5.

shooting more 3-pointers as well, meaning the results are consistent with the offensive having a greater ability to adjust than the defense (qualitatively consistent with the no defensive adjustment model). The offense shoots more 3's and the average value falls—making this trade-off respects the true value of 3-pointers in terms of winning the game and is consistent with a downward sloping usage curve.

The results take a different turn when we examine the behavior of the leading team. β_4 is estimated to be significantly *positive* (about 1/3 the magnitude of β_5). This is the wrong direction and is a violation of optimal shot selection. Recall that we found in Table 1 that the leading team tends to shoot fewer 3's when α increases. Here we see that this decrease in usage is accompanied by an increase in efficiency, which is again consistent with a downward sloping usage curve and limited defensive adjustment. For a leading team, as the game gets closer, the team *should* become more risk-neutral, yet the team actually behaves in a more risk averse manner. Leading teams do not appear to place the right price on risk, whereas trailing teams do.

4.3 The Rubber-band effect and performance in the clutch

So far we have documented two important behavioral patterns that increase the likelihood of a comeback by the trailing team. First, the trailing team shows a boost in efficiency for both shot types. Second, the leading team tightens up as the score gets closer – shooting fewer 3's and more 2's, contrary to the change in their incentives. We call the combined impact the "rubber band effect."[6] More games tend to be close late than we'd otherwise expect to observe. What this means is that performance in the clutch tends to be very important in the NBA, because clutch moments are relatively frequent.

The natural question is "how is team quality related to clutch performance." To answer this question we estimate how gross offensive efficiency relates to the win value of a point (a natural measure of how importance points are). Our estimating equation is given by, where we again use line-up fixed effects:

$$E(\text{Points}_p) = \delta_{Off_p} + \gamma_{Def_p} + \beta_{1,oTeam_p} \cdot WVP_p - \beta_{2,dTeam_p} \cdot WVP_p$$

We plot the results in Figures 2 and 3. The y-axis measures the difference in performance in a clutch moment (WVP=0.07, a moment in which each point increases the teams chance of winning by 8 percentage points, which corresponds to about the 97th percentile of importance) as compared to how that team does in a median moment (WVP=0.02). We call this measure the "clutch bonus" and plot it as points per 100 possessions. In Figure 2, the x-axis gives baseline offensive efficiency (points scored per 100 possessions in typical situations). This is a natural measure of pace-adjusted offensive quality. Each dot represents a team-year. It is clear that on average teams do

[6] Substitutions to conserve star players for games down the road could also contribute and is outside the scope of our 'fixed lineup' analysis.

worse in clutch moments—for most teams it's harder to score when the chips are down. The slope of the fitted line is 1.65 and highly significant ($t = 7.27$). This indicates that the better a team is at offense, the better it does in the clutch relative to its own baseline—bad teams do much than their worse than their baseline, average teams do a little worse and very good teams actually get better in the clutch.

In Figure 3 shows how defense quality relates to clutch performance. Defensive efficiency, given on the x-axis, is points allowed per 100 possessions, so lower numbers correspond to stingier defenses. The y-axis gives the clutch bonus of the opposing team. It measures how well the defense performs in the clutch, with negative values being good. Again we see most values are negative—the average defense is better in the clutch. The clutch bonus is more negative for better defensive teams; for rather poor defensive teams the value is positive, meaning they consistently do worse in the clutch as compared to baseline scenarios. The slope of the fitted line is -1.09 ($t = 4.20$).

Comparing the absolute values of the fitted lines and the R^2 of the regressions, we conclude the impact of unit quality on clutch performance is significantly stronger and less noisy on the offensive side of the ball. What this means is that for two evenly matched-up teams in terms of baseline performance, a team with a good offense will tend to have an advantage in the clutch. For teams of differing overall ability, the better team will have an advantage in the clutch on both sides of the ball and will thus tend to pull out more close games than their baseline advantage would suggest.

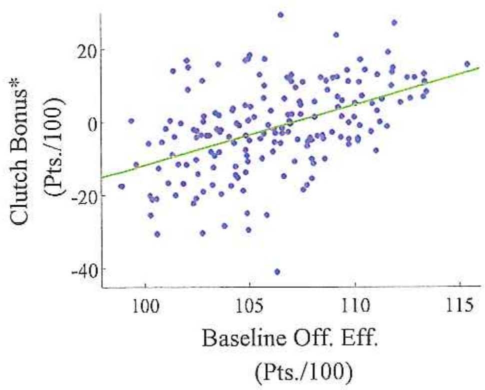


Figure 2: Offense clutch performance vs. baseline efficiency.

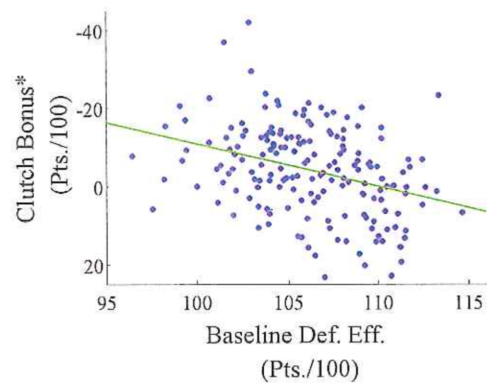


Figure 3: Defense clutch peformance vs. baseline efficiency.

5 Conclusion

We theoretically and empirically investigate the optimality of NBA shooting decisions in response to changing incentives over risk. The most robust theoretical requirement is that the gap in nominal efficiency between 3 and 2-point attempts must be negatively correlated with the offense's preference for risk. We find adherence to our key test of optimality for the trailing team—as a trailing team gets in a more desperate situation (becomes more risk-loving), the efficiency of their 3-point attempts falls. The trailing team also tends to increase their fraction of 3-point attempts in proportion to their preference for risk, consistent with the ability to shift the offensive attack more flexibly than the defense can adjust resources. The leading team, however, violates our optimality requirement; leading teams shoot fewer 3's as their preference for risk increases and these 3's actually offer higher average point value (consistent with a downward sloping usage curve). As a lead decreases, the leading team should become more risk-neutral, but teams in this circumstance actually tighten up and become more risk averse, contrary to what their risk preferences ought to be to maximize the chance of winning the game.

We also find a strong motivational effect of trailing on the scoreboard—a given lineup sees a boost inefficiency for both 2's and 3's—combined with the sub-optimal shot selection of the leading teams this helps explain the surprising frequency of comebacks in the NBA and means clutch moments tend to occur more frequently that we'd otherwise expect. We show that for an average team it's harder to score in clutch moments, but very good offenses actually do better in the clutch, whereas bad defenses actually get worse. Taken together, this means good teams have an even greater advantage when the chips are down.

References

[1] J. Berger and D. Pope. Can losing lead to winning? *Management Science*, 57(5):817–827, 2011.

[2] Matthew Goldman and Justin M. Rao. Allocative and dynamic efficiency in nba decision making. In *Proceedings of the MIT Sloan Sports Analytics Conference*, pages 1–10, 2011.

[3] Matthew Goldman and Justin M. Rao. Effort vs. concentration: The asymmetric impact of pressure on nba performance. In *Proceedings of the MIT Sloan Sports Analytics Conference*, pages 1–10, 2012.

[4] Matthew Goldman and Justin M. Rao. Motivational loss aversion. 2013.

[5] D. Oliver. *Basketball on paper: Rules and tools for performance analysis*. Brassey's Press, 2004.

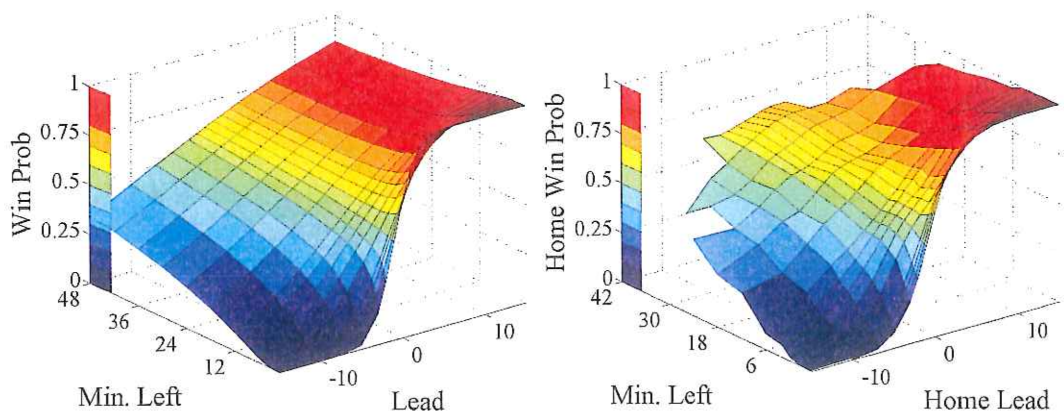
[6] S. Rick and G. Loewenstein. Hypermotivation. *Journal of Marketing Research*, 45(6):645–648, 2008.

[7] B. Skinner. Scoring strategies for the underdog: Using risk as an ally in determining optimal sports strategies. *Journal of Quantitative Analysis in Sports*, 7, 2011.

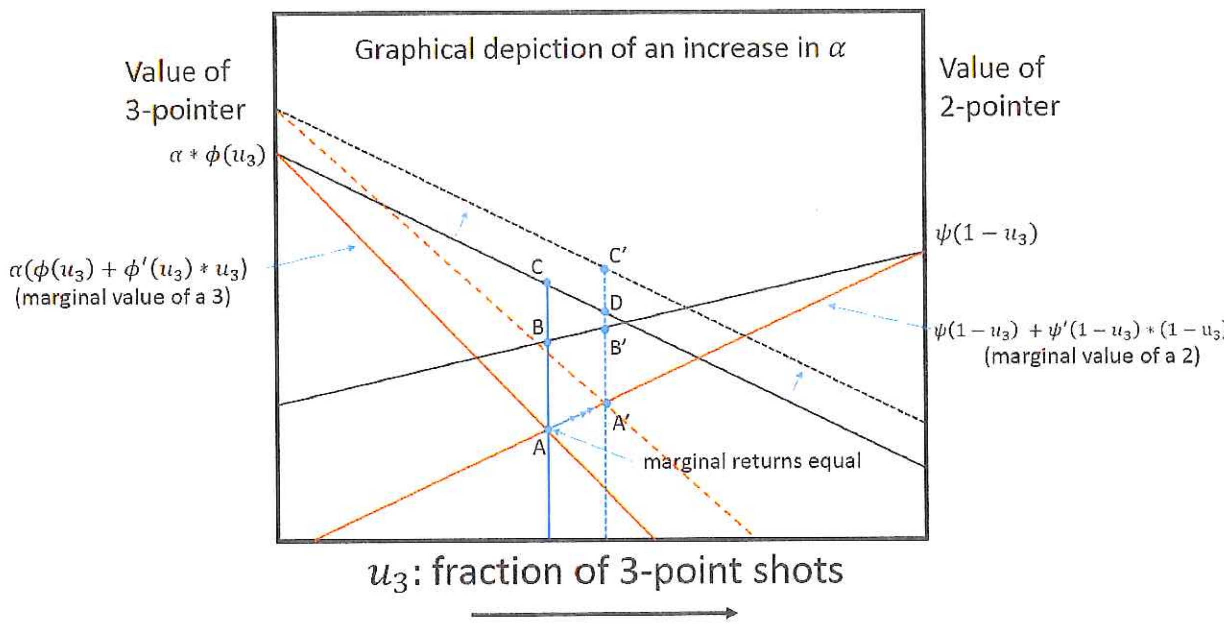
[8] A. Tversky and D. Kahneman. Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics*, 106(4):1039–1061, 1991.

6 Appendix

6.1 Figures



Appendix Figure 1: Parametric projects of win probability conditional on score margin and time remaining for the home team in even match-up; Panel 2: non-parametric estimates of the same function.



Appendix Figure 2: Graphical representation of the no-defensive adjustment model. The initial "profit maximizing" condition given by the line intersecting point A and the impact of an increase in α with the new equilbrium given by the line intersecting A'.

6.2 The model with defensive adjustment

Offense (defense) seeks to maximize (minimize) the offenses increase in win probability in a given possession. This utility function (for the offense) is

$$U = u_3 p_3 WV_3 + u_2 p_2 WV_2$$
$$\frac{U}{WV_2} = \alpha u_3 p_3 + u_2 p_2$$

subject to the constraints that

$$u_2 = 1 - u_3$$
$$d_2 = 1 - d_3$$
$$p_3 = \phi(u_3, d_3)$$
$$p_2 = \psi(u_2, d_2).$$

We assume the following (written in terms of ϕ but they all hold for ψ too):

1. Usage curves are downward sloping: $\phi_1 < 0$.

2. Usage curves are such that marginal shots have declining value: $\frac{d^2(u_3 \cdot \phi(u_3))}{du_3^2} = 2\phi'(u_3) + u_3\phi''(u_3) < 0$

3. Defensive pressure lowers shooting percentage: $\phi_2, \psi_2 < 0$.

4. Defense has diminishing returns: $\phi_{22}, \psi_{22} > 0$.

5. Using more possessions in a certain way increases (makes more negative) returns to defense against that type of use: $(\phi_2 + u_3^* \phi_{21}) < 0$.

Let starred values denote the equilibrium quantities. Then the defense's first order condition is given by

$$\alpha u_3^* \phi_2(u_3^*, d_3^*) = (1 - u_3^*)\psi_2(1 - u_3^*, 1 - d_3^*), \tag{4}$$

where the subscript denotes a derivative in the corresponding argument. The offense's first order condition is given by

$$\alpha[\phi(u_3^*, d_3^*) + u_3^* \phi_1(u_3^*, d_3^*)] = [\psi(1 - u_3^*, 1 - d_3^*) + (1 - u_3^*)\psi_1(1 - u_3^*, 1 - d_3^*)] \tag{5}$$

where the bracketed quantities represent marginal shot probabilities for 3 and 2 point shots respectively. Both of these must both be greater than 0. Taking total differentiation of (4) and omitting the arguments of ϕ and ψ yields

$$u_3^* \phi_2 d\alpha + \alpha u_3^* \phi_{22} dd_3^* + \alpha(\phi_2 + u_3^* \phi_{21}) du_3^* = -u_2^* \psi_{22} dd_3^* - (\psi_2 + u_2^* \psi_{21}) du_3^*$$

and rearranges to

$$\begin{aligned} u_3^* \phi_2 d\alpha + [\alpha(\phi_2 + u_3^* \phi_{21}) + (\psi_2 + u_2^* \psi_{21})]\, du_3^* + [\alpha u_3^* \phi_{22} + u_2^* \psi_{22}] dd_3^* &= 0 \\ b_2 d\alpha + a_{21} du_3^* + a_{22} dd_3^* &\equiv 0, \end{aligned} \tag{6}$$

where the values of a_{11}, a_{12} and b_1 are defined implicitly. A similar analysis of equation (5) gives

$$\begin{aligned} [\phi(u_3^*, d_3^*) + u_3^* \phi_1(u_3^*, d_3^*)] d\alpha + [(2\phi_1 + u_3^* \phi_{11}) + (2\psi_1 + u_2^* \psi_{11})] du_3^* & \\ + [\alpha\,(\phi_2 + u_3^* \phi_{12}) + (\psi_2 + u_2^* \psi_{12})]\, dd_3^* &= 0 \\ b_1 d\alpha + a_{11} du_3^* + a_{12} dd_3^* &\equiv 0. \end{aligned} \tag{7}$$

where the values of a_{21}, a_{22} and b_2 defined implicitly. In matrix notation

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} du_3^* \\ dd_3^* \end{bmatrix} = - \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} d\alpha$$

Then by Cramer's Rule

$$\frac{dd_3^*}{d\alpha} = - \frac{\begin{vmatrix} a_{11} & b_1 \\ a_{21} & b_2 \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}} = \frac{\overbrace{a_{21} b_1}^{(-)} - \overbrace{a_{11} b_2}^{(+)}}{\underbrace{a_{11} a_{22} - a_{12} a_{21}}_{(-)}} > 0$$

To sign these derivatives not that, $a_{11} < 0$ (marginal shots have diminishing value), $a_{22} > 0$ (diminishing returns to defense), and $a_{12} = a_{21} < 0$ (shooting more 3s raises the effectiveness of defense on 3s). Thus both denominators are negative. Also $b_1 > 0$ (its a marginal shot value) and $b_2 < 0$ ($\phi_2 < 0$). Signing this derivative states that defensive pressure on 3's must increase with α.

Proof of Proposition 2

$$\frac{du_3^*}{d\alpha} = -\frac{\begin{vmatrix} b_1 & a_{12} \\ b_2 & a_{22} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}} = \frac{\overbrace{b_2 a_{12}}^{(+)} - \overbrace{b_1 a_{22}}^{(+)}}{\underbrace{a_{11}a_{22} - a_{21}a_{12}}_{(-)}} = ?.$$

Manipulating the numerator, we have that $\frac{du_3^*}{d\alpha} >$ if and only iff:

$$\frac{a_{12}}{b_1} > \frac{a_2 2}{b_2}$$

We first note that both sides of this inequality are negative, it is convenient to write:

$$\left|\frac{a_{12}}{b_1}\right| < \left|\frac{a_{22}}{b_2}\right|$$

$a_{12} = \alpha(\phi_2 + u_3^*\phi_{12}) + (\psi_2 + u_2^*\psi_{12})$ is the cross-partial marginal effect of defense. It says "how much more effective does defense become when an offense increases its fraction of 3's. This terms gives the incentive for the defense to adjust into 3's. b_1 is the offense's marginal shot value of a 3, as the usage curve gets steeper, this value falls. On the RHS, the numerator is a term, $\alpha u_3^*\phi_{22} + u_2^*\psi_{22}$, that captures the concavity of the defense's response function. The denominator captures the marginal impact of defense. If the above equation holds, the offense will take more 3's when their preference for risk increases. This equation says this is more likely to occur when the defense has a concave adjustment function (they face strong diminishing returns to selective pressure), when the cross partial is low (the extra impact of guarding 3's does not increase much with the offense's 3-point usage) and when the usage curve of a 3-pointer relatively flat (raising the marginal value of a 3-pointer, which raises the denominator on the LHS).

Proof of Proposition 3

Other comparative statics are directly implied by our constraints,

$$\frac{du_2^*}{d\alpha} = -\frac{du_3^*}{d\alpha} = ?$$
$$\frac{dd_2^*}{d\alpha} = -\frac{dd_3^*}{d\alpha} < 0$$
$$\frac{dp_3^*}{d\alpha} = \frac{du_3^*}{d\alpha}\phi_1 + \frac{dd_3^*}{d\alpha}\phi_2$$
$$= -\frac{\phi_1(\overbrace{b_1 a_{22}}^{(+)} - \overbrace{b_2 a_{12}}^{(+)}) + \phi_2(\overbrace{a_{11}b_2}^{(+)} - \overbrace{a_{21}b_1}^{(-)})}{\underbrace{a_{11}a_{22} - a_{12}a_{21}}_{(-)}}$$

which first does not appear signable, but can be rearranged to

$$= -\frac{\overbrace{\phi_1 b_1 a_{22}}^{(-)} + \overbrace{b_2(\phi_2 a_{11} - \phi_1 a_{12})}^{(-)} - \overbrace{\phi_2 a_{21} b_1}^{(+)}}{\underbrace{a_{11}a_{22} - a_{12}a_{21}}_{(-)}} < 0,$$

where the middle term in the numerator can be signed by noting that $b_2 < 0$ and $(\phi_2 a_{11} - \phi_1 a_{12}) = \phi_1\phi_2(1 + 2u_3^*) > 0$.

$$\frac{dp_2^*}{d\alpha} = \frac{du_2^*}{d\alpha}\psi_1 + \frac{dd_2^*}{d\alpha}\psi_2 < 0,$$

follows by symmetry to the above calculation.

6.3 Proofs for the baseline model

Proof of Proposition 1 The only part of Proposition 1 not shown in the text is that the win value of 3's must increase. We think intuition can be best scene through the lense of a classic economics setup. Consider a monopolist facing demand curve $P(q)$ and an upward slope marginal cost curve $C''(q) < 0$. Imagine a subsidy from the government of so that for each dollar earned, the firm earns $1 + x = \alpha > 1$ dollars. What the proposition states is that if the government offers subsidy x, the price cannot fall by more than x.

This problem is isomorphic to our shot allocation problem because the downward sloping 2-point usage curve implies an increasing marginal opportunity cost of shooting 3's. As I shoot more 3's, I give up better and better 2-pointers. The first order condition of this problem is:

$$\alpha MR(q) = MC(q)$$

Taking the total derivative, rearranging and multiplying by $\frac{dp}{dq}$ we get:

$$\frac{dp}{d\alpha} = \left(\frac{MR}{MC' - \alpha MR's}\right)\frac{dp}{dq}$$

We are interested in whether $p * \alpha$ is greater than the orginal price, this amounts to whether:

$$\frac{d(p\alpha)}{d\alpha} = \alpha * \frac{dp}{d\alpha} + p > 0$$

Plugging, in our condition becomes, is:

$$p = \left(\frac{\alpha MR}{\alpha MR's - MC'}\right)\frac{dp}{dq}$$

$$p > \frac{(p + p'(q)q)p'(q)}{p''(q)q + 2p'(q)}$$

Cross-multiplying and rearranging we have:

$$qp''(q) < \frac{q(p'(q)^2 q - p'(q)p)}{pq}$$

$$-2p'(q) < \frac{q(p'(q)^2 q - p'(q)p)}{pq}$$

where the second line follows because $qp''(q) + 2p'(q) < 0$ (marginal revenue is downward sloping). Canceling out and simplying, this equation reduces to:

$$p'(q) > -\frac{p}{q}$$

$$p'(q) * \frac{q}{p} > -1$$

$$\frac{1}{\epsilon} > -1$$

where ϵ is the elasticity of demand. The last line must hold, otherwise the firm earns negative marginal revenue.

6.4 Parametric model of win probability

A game of NBA basketball has 48 minutes of game time, with ties being settled by a 5-minute overtime. Consider two teams, home (h) and away (a). Let $S_{h,N}$ and $S_{a,N}$ denote the current scores for the home and away team with N offensive possessions (for each team) remaining in the game. Let $P_{h,i}$ and $P_{a,i}$ denote the number of points scored by the home/away team on the i^{th} possession from the end of the game. The home team wins if it has more points at the end of the game, which we can express as:

$$S_{h,0} > S_{a,0} \iff S_{h,N} + \sum_{i=1}^{N} P_{h,i} > S_{a,N} + \sum_{i=1}^{N} P_{a,i} \iff \sum_{i=1}^{N} P_{h,i} - P_{a,i} > S_{a,N} - S_{h,N}.$$

To model how teams generate points, let $\{\mu_h, \sigma_h^2\}$ and $\{\mu_a, \sigma_a^2\}$ represent the mean and variance of points per possession that each team is able to achieve in the match-up. If the number of remaining possessions, N, is large, the central limit theorem gives the probability of the home team winning as:

$$\begin{aligned} P(\text{Home Win}) = P(S_{h,0} > S_{a,0}) &= P\left(\sum_{i=1}^{N}(P_{h,i} - P_{a,i}) > S_{a,N} - S_{h,N}\right) \\ &= \Phi\left(\frac{S_{h,N} - S_{a,N} + N(\mu_h - \mu_a)}{\sqrt{N(\sigma_h^2 + \sigma_a^2)}}\right), \end{aligned} \tag{8}$$

where Φ is the CDF of the standard normal distribution. Examining this expression, we see that an ability advantage (μ higher than opponent) matters proportional to the number of remaining possessions. Each factor's marginal impact on winning the game is easily obtained by differentiating equation (8). The following expression gives the impact of a point scored for the home team on win probability:

$$\frac{dP(\text{Home Win})}{dS_{h,N}} = \phi\left(\frac{(S_{h,N} - S_{a,N}) + (N(\mu_h - \mu_a))}{\sqrt{N(\sigma_h^2 + \sigma_a^2)}}\right)\frac{1}{\sqrt{N(\sigma_h^2 + \sigma_a^2)}}, \tag{9}$$

where lower-case ϕ is the standard normal PDF. To estimate this equation, we first impute the number of remaining possessions using the team-specific paces-of-play in a given match-up and by adding one possession to the team currently holding the ball. Given the standard normal specification, it is natural to estimate equation (8) with Probit regression. The projections give the probability the home team will win at each of state of the game. Figure 1 Panel 1 shows these projections.

The Dwight Effect: A New Ensemble of Interior Defense Analytics for the NBA

Kirk Goldsberry[1] and Eric Weiss[2]
[1]Harvard University,
Cambridge, MA, USA, 02138
kgoldsberry@fas.harvard.edu

[2]Sports Aptitude, LLC
eweiss@sportsaptitude.com

Abstract
Basketball is a dualistic sport: all players compete on both offense and defense, and the core strategies of basketball revolve around scoring points on offense and preventing points on defense. However, conventional basketball statistics emphasize offensive performance much more than defensive performance. In the basketball analytics community, we do not have enough metrics and analytical frameworks to effectively characterize defensive play. However, although measuring defense has traditionally been difficult, new player tracking data are presenting new opportunities to understand defensive basketball. This paper introduces new spatial and visual analytics capable of assessing and characterizing the nature of interior defense in the NBA. We present two case studies that each focus on a different component of defensive play. Our results suggest that the integration of spatial approaches and player tracking data promise to improve the status quo of defensive analytics but also reveal some important challenges associated with evaluating defense.

Introduction
Basketball is a dualistic sport. Players compete on both offense and defense, and the two core objectives of all basketball stratagems are scoring points and preventing points. Although it is self-evident that the final score of every basketball game depends equally on these two facets, this basic tenet is not properly represented in contemporary basketball statistics. A quick reading of even the most "advanced" basketball statistics would suggest that basketball success hinges more on offensive factors and less on defensive factors. Few of the sport's most common metrics quantify key defensive aspects. Basketball's most common statistics are related to events that are most obviously attributable to one individual action at one moment; defensive prowess in basketball fails to meet this basic criterion.

Contemporary basketball expertise is significantly hindered by the inability to properly assess defensive play; current evaluations of a player or team's defensive tendencies are constrained by a lack of proper reasoning artifacts. Most defensive analytics remain guided by the simple tallying of disparate event types including "steals," "blocks," and "defensive rebounds," which does little to characterize either the nature or the effectiveness of defensive performance. Effective defensive play requires a cohesive assembly of structured actions converging upon a simple objective: keep your opponent from scoring points. With this in mind, as the NBA enters its "big data" era and new kinds of basketball analytics emerge, advancing defensive understanding presents one of our biggest challenges.

This paper explores defensive evaluations in the NBA and examines emerging opportunities and challenges associated with measuring defense using optical tracking data. The paper presents a new methodology designed to characterize the interior defensive effectiveness of NBA "big men". The core objectives of this paper are 1) to improve the characterization and understanding of interior defense in the NBA, and 2) expose key challenges associated with measuring defense as new forms of performance data emerge. We present case studies that 1) use spatial analyses to extract new defensive metrics from optically tracked game data (SportVu data) and 2) use visual analytics to present results.

The paper also introduces a new ensemble of spatially minded metrics that present a novel and simple means to characterize basketball performance. One key and recurring limitation of many basketball statistics is their relatively limited explanatory abilities. For example, even the most effective "advanced" metrics like "defensive rating" (points allowed per 100 possessions) may provide valuable insight into overall performance ability, but simultaneously they often fail to offer any additional explanatory insight as to *why* a performance may be good or bad. We introduce "spatial splits" - a concept inspired by baseball's "triple-slash" lines - as a means to address this shortcoming; in tandem with other metrics, we contend spatial splits provide additional insight into the nature of how players and teams are performing within court space, therein providing analysts with a more powerful set of reasoning artifacts.

The paper contains three main sections: a brief background section is followed by an explanation of our methodology, which in turn is followed by a discussion of our results and conclusions. We also append thorough listings of detailed results at the end of the paper.

Background

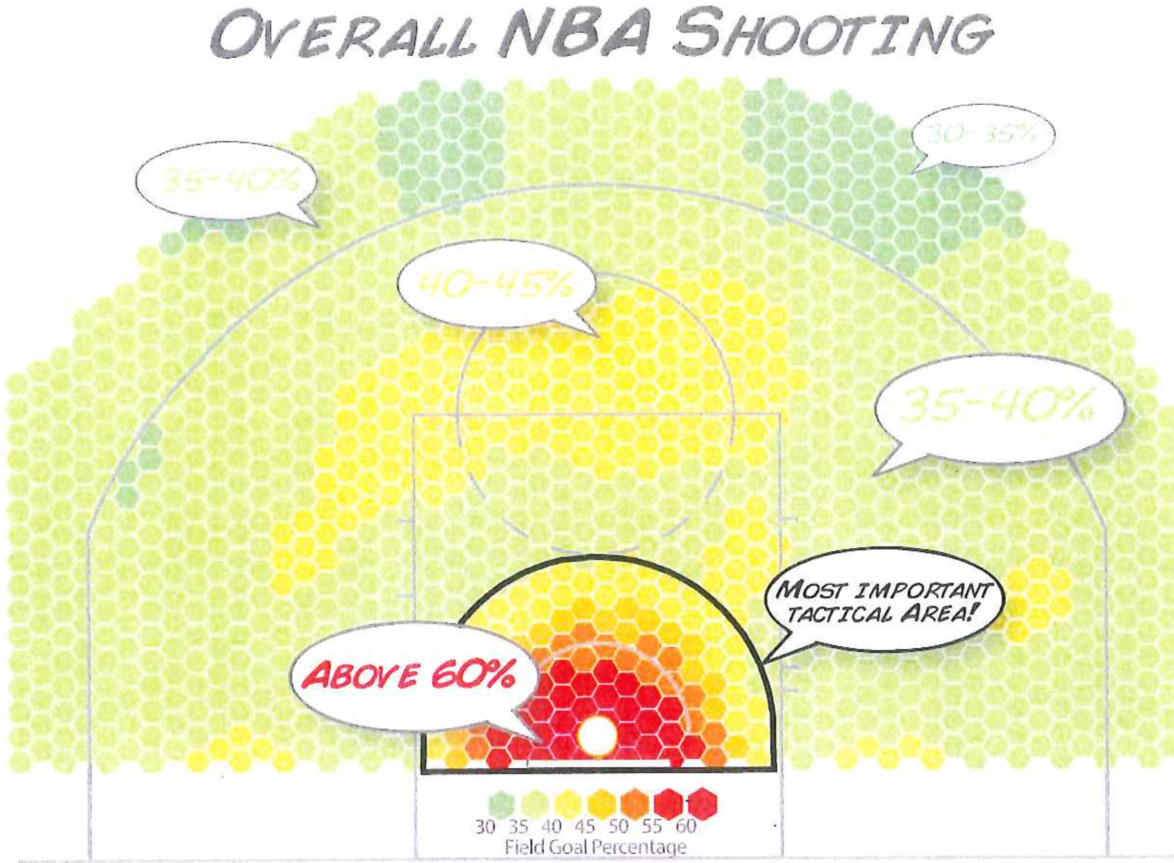


Figure 1: Overall shooting efficiencies in the NBA. The only shots that go in over half the time occur close to the basket. For this reason, this relatively small area remains the most important tactical space - and the most vigorously defended space in the NBA. Graphic by Kirk Goldsberry.

NBA shooters only make about 39% of their field goals from everywhere outside of 7 feet. The only shots that go in more than half the time occur very close to the rim. Despite the rapidly growing importance of the 3-point shot, good shots close to the basket remain the best shots on the floor; not only do they result in points at a higher rate, when missed they have a much greater chance of being rebounded by the shooting team. Over 70% of shots near the rim either result in points, a shooting foul or an offensive rebound. Good shots near the rim are clearly advantageous. For this reason, the league shoots over 1/3rd of its shots from the tiny portion of the court close to the basket, and defenders protect this area with more vigor than any other real estate on the court. Although the vitality of this strategic space is self-evident, few if any contemporary analytics effectively characterize the ability of players or teams to defend basketball's most sacred real estate. The problem is obvious: interior defense is critical to basketball success, but our ability to measure or characterize players' interior defensive abilities remains undeveloped. Consider these two basic questions:

1) Who is the best interior defender in the NBA?
2) What metrics would you use to answer that question?

The NBA's most prominent defensive metrics can be misleading, but this is not a problem unique to basketball. Until very recently, the dominant conventional defensive metrics in baseball were "errors" and "fielding percentage," which do not frequently correlate with a player's true defensive value. In the NFL, the best cornerbacks never lead the league in any conventional stats because quarterbacks are too afraid to even throw in their direction; they don't even get chances to defend passes. Basketball exhibits similar issues; our conventional defensive metrics fail to accurately reveal the NBA's most dominant defenders.

Last season, Oklahoma City's Serge Ibaka led the NBA in blocks by averaging an incredible 6.46 blocks per 48 minutes, but what does that really reveal? Does that mean he is an "elite defender," or even the "best shot blocker" in the NBA? Shot blocks are relatively infrequent events that have an ambiguous relationship with defensive effectiveness. In many cases, for a shot block event to occur a shooter has to believe that his shot will not be blocked. In other words, the shot blocker has to "come out of nowhere" or has to somehow deceive the shooter; at the point of the shot's release the shooter believes the path is clear, but that turns out not to be the case.

Dwight Howard, who is commonly referred to as the NBA's "most dominant" interior defender, only averaged 2.69 blocks per 48 minutes, almost 4 fewer than Ibaka; however, it could be argued that Howard's mere presence "blocks" shots before they happen. The presence of a truly dominant interior force can augment the spatial behavior of the offense in the same way that a dominant cornerback changes the behavior of a quarterback. While it is easy to tally up things like blocks, rebounds, and steals, it's much harder to measure the kind of disruption or the strategic augmentations that dominant interior defenders like Dwight Howard create. We define "The Dwight Effect" as the ability of an interior defender to reduce the efficiency of an opponent's shooting behavior.

Perhaps the most logical method to evaluate this disruption is to measure the spatial shooting patterns and efficiencies of NBA teams in the presence of different interior defenders. Using emerging data sets from SportVu, it's now possible – although still not easy – to look at defense in new ways. In the case of interior defense, we can evaluate how NBA offenses behave differently depending on which NBA "bigs" are on the floor; furthermore, we can evaluate how offenses behave when a given NBA interior defender is "protecting the rim" or near a shot event.

Methodology, Data, and Case Studies

We conducted two separate case studies of interior defense in the NBA. Using player tracking data provided by STATS (SportVu) we evaluated player positions, shooting tendencies, and shot outcomes for over 75,000 NBA shots during the 2011-2012 and 2012-2013 seasons. We evaluated the spatial structures and efficiencies of NBA shooting in the presence of the 52 NBA interior defenders who faced at least 500 shot attempts during the study period. Each case study monitors a different aspect of defensive effectiveness and introduces new metrics.

We introduce "spatial splits" as a means to communicate our results. Since NBA scoring efficiency is clearly dependent on spatial factors, we contend spatial splits offer a mechanism to detect, understand, and communicate key aspects of NBA scoring efficiency. Presented in a manner meant to mimic baseball's "slash line" or "triple-slash line" these sequences of three numbers not only offer a basic quantification of a player or opponent's shooting, they also present an inherent explanatory characterization as well. Figure 2 depicts the 3 zones represented in the spatial splits.

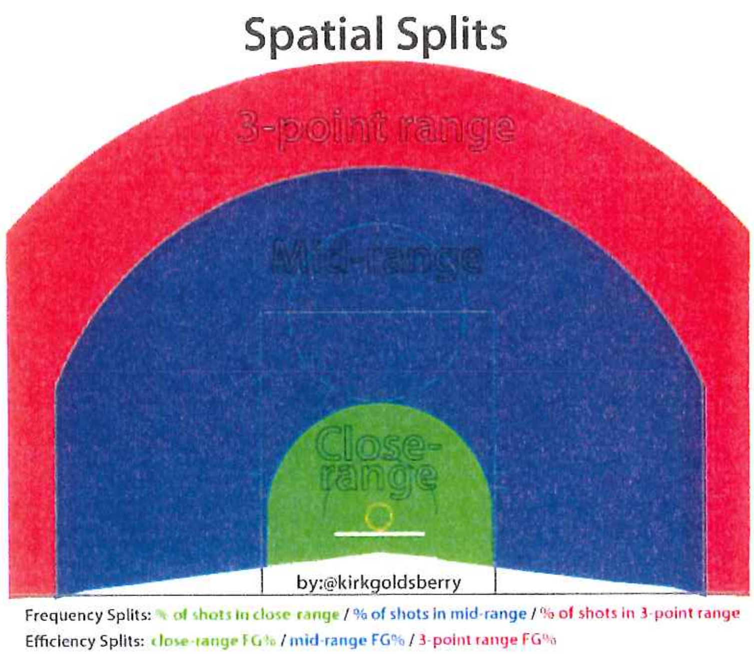


Figure 2: The 3 zones associated with spatial splits: close-range in green, mid-range in blue, and 3-point range in red. Splitting offensive performance data using these zones can help characterize the nature of scoring behaviors in the NBA.

We introduce two kinds of spatial splits: frequency splits and efficiency splits. Both reflect percentage values in the following sequence: Close-range value / Mid-range value / 3-point range value. Frequency splits focus on shot distribution; each number corresponds to the percentage of shots that come from the corresponding zone. The three numbers in the frequency splits should sum to 100 (barring any rounding errors). Efficiency splits characterize how well a player or teams shoots from each zone; each value represents the field goal percentage in the corresponding zone. As an example, consider the spatial splits of the NBA as a whole, and two NBA players from the previous two NBA seasons: Kevin Durant and Josh Smith.

NBA League Average: Frequency 35/41/24 Efficiency 53/39/36
Kevin Durant: Frequency 27/46/27 Efficiency 65/44/37
Josh Smith: Frequency 45/43/12 Efficiency 61/37/30

The above examples illustrate the ability of spatial splits to quickly summarize key differences in scoring tendencies. These splits quickly communicate a few facts: 1) In terms of shot distribution, Kevin Durant shot 27% of his shots close to the basket, 46% of his shots in the mid-range, and 27% of his shots from three-point range, 2) In terms of shot efficiency, Durant shot 65% close to the basket, 44% in the mid-range, and 37% from beyond the arc – above league averages in each zone, while Smith is only above average close to the basket. They also enable comparison across players. In this case we can quickly note that Durant is less active close to the basket than Smith, they are both active in the midrange, Durant is more active beyond the arc, and Durant is a more efficient shooter in every area. We contend that this contribution is a valuable new way to characterize NBA scoring behaviors.

Although spatial splits present an effective way to characterize the nature of an individual player's offensive tendencies and abilities, in this paper we use them to evaluate defense. More specifically, within the context of spatial splits, effective interior defense should manifest in two ways. The most obvious is perhaps reduced shooting *efficiencies* close to the basket. The second is less apparent but perhaps more important: reduced shooting *frequencies* close to the basket, and increased frequency in the mid-range and three-point areas. Taken together, reduced close-range efficiency and reduced close-range frequency translate to fewer easy shots, fewer points, and fewer offensive rebounding opportunities for the offense.

Case Study 1: The Basket Proximity Condition

The objective of the first case study was to examine the ability of interior defenders to "protect the basket." This case study considered shot attempts that occurred when there was an interior defender within 5 feet of the basket and was designed to measure two aspects of point prevention: the ability to prevent shots near the basket, and the ability to reduce the shooting efficiency of opponents near the basket. We evaluated shooting patterns using spatial splits. As a means to characterize the opponents' shooting tendencies, we calculated both the frequency and efficiency of shooting in each zone, but placed primary emphasis on close range shooting.

Case Study 2: The Shot Proximity Condition

The second case study evaluates the ability of interior defenders to defend shots in their immediate proximity. This study has two objectives: to determine how frequently an interior defender is proximate to a shot attempt, and to determine how effective an interior defender is when they are proximate to a shot attempt. In this case we place a reduced emphasis on shot locations and instead evaluate two other aspects of defending; each aspect is evaluated via a new metric:

A) Shots Defended: the relative frequencies in which the defender finds himself within 1, 3, or 5 feet of shot attempts.
B) Proximal FG%: the relative efficiencies of shooters in the proximity of the defender.

Results

Case Study 1: Basket Proximity

Overall more than 1/3rd of shots in our superset of 76,000 shots occurred with an interior defender within 5 feet of the basket. We assert that "dominant" interior defense can manifest in two ways: reducing the shooting efficiency of opponents, and also reducing the shooting frequency of opponents. In terms of reducing efficiency, we found that Indiana's Roy Hibbert and Milwaukee's Larry Sanders (Figure 3) were by far the most effective. We evaluated this by measuring the field goal percentage of close range shots when a qualifying interior defender was within 5 feet of the basket. Overall, NBA shooters make 49.7% of their field goal attempts when qualifying interior defender is within 5 feet of the basket; however, this number drops to 38% when either Hibbert or Sanders are within 5 feet. In contrast, we found that Phoenix's Luis Scola and Golden State's David Lee (Figure 3) were the worst defenders in these situations; opponents made 63% of their close-range field goals when Scola was within 5 feet of the basket. See Appendices 1 and 1A for a full list of qualifying defenders.

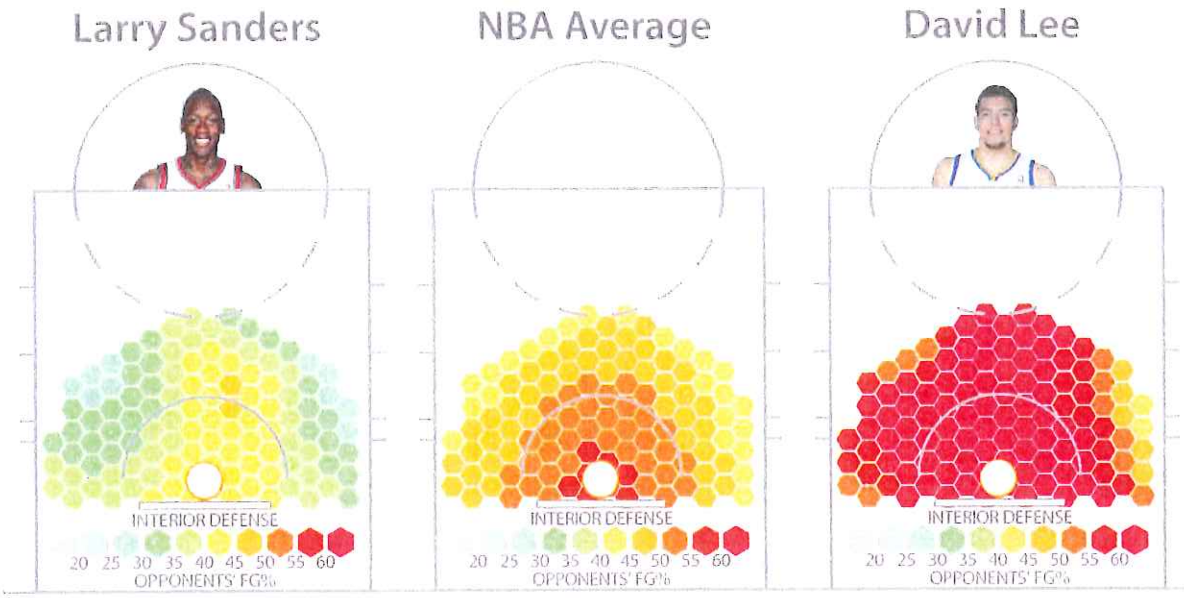


Figure 3: Opponents' field goal percentages vary widely depending on which interior defender is close to the basket. Milwaukee's Larry Sanders is one of the most effective interior defenders in the league; opponents struggle to score when he is near the basket. This is not the case with Golden State's David Lee; when he is close to the basket, opponents score at very high efficiencies.

We also contend that dominant interior defenders often deter shots from even happening. Many NBA players will be reluctant to "challenge" a dominant interior player or be more likely to "settle" for a jump shot further from the basket. We evaluated this effect by examining the percentage of field goal attempts that occur near the basket when a qualifying interior defender is within 5 feet of the rim. We found that the most deterrent interior defender in this sense was Dwight Howard. Overall, when a qualifying defender is within 5 feet of the basket, the NBA shoots 57.2% of its attempts close to the basket; however, when Dwight Howard was the interior defender this number dropped to 48.2% (Appendix 1A). This is what we call the "Dwight Effect" – the most effective way to defend close range shots is to prevent them from even happening. Although Howard does not lead the league in blocks, he does lead the league in "invisible blocks," which may prove to be markedly more significant. When Howard is protecting the basket, opponents shoot many fewer close range shots than average, and settle for many more mid-range shots, which are the least productive shots in the NBA. Furthermore, out of centers who have faced at least 100 total shots in the basket proximity study, Serge Ibaka ranked last; when he is within 5 feet of the basket, opponents shot 74% of their shots in the close range area. This means that Ibaka is likely to be around any shot near the basket and suggests that while Ibaka leads the NBA in blocks per game, part of the reason is that he has many more "potential blocks" than almost any other defender. Full results for Case Study 1 are presented in Appendix 1. We also discuss the challenges and limitations associated with this study in the Discussion section.

Case Study 2: Shot Proximity
Overall 27.8% of NBA shots occur within 5 feet of a qualifying interior defender. We evaluated 21,042 shots that met this criterion and examined two separate aspects of defensive tendencies: Shots Defended, and Proximal FG%. The results for each are presented below.

A) Shots Defended: the relative frequencies in which the defender is located within 1, 3, or 5 feet of shot attempts. We calculated these frequencies for 93 qualifying defenders that faced at least 200 shots while playing defense. Tyler Hansbrough had the lowest frequency of being close to shots; he was within 5 feet of shots only 20.7% of the time. Only 3 defenders were within 5 feet of shots more than 35% of the time: Josh Harrellson, Kosta Koufos, and Jordan Hill. Serge Ibaka was fourth at 34.5%. The full set of results is available in Appendix 2.

B) Proximal FG%: the relative efficiencies of shooters in the proximity of the defender. Overall, when there is a qualifying interior defender within 5 feet of a shot attempt, the NBA shoots 45.6% from the field; however this value varies considerably depending on which defender that is. The most effective proximate defender in our study was Larry Sanders; opponents shot only 34.9% when he was within 5 feet of their shot. Conversely, Anderson Varejao was found to be the least effective proximate defenders with a proximal FG% value of 54.2%. Table 1 summarizes the best and worst players according to proximal FG%, but a complete list of proximal FG% values can be found in Appendix 2.

Top 5	Proximal FG%
1. Larry Sanders	34.9%
2. Andrea Bargnani	35.2%
3. Kendrick Perkins	37.3%
4. Elton Brand	38.0%
5. Roy Hibbert	38.7%

Bottom 5	Proximal FG%
48. Kevin Love	52.1%
49. Jonas Valanciunas	52.8%
50. David Lee	53.0%
51. Jordan Hill	53.9%
52. Anderson Varejao	54.2%

Table 1: The top and bottom 5 interior defenders according to proximal FG%, which is defined as the opponent's FG% when the qualifying defender is within 5 feet of the shot attempt.

Discussion and Limitations

In a league that is both teeming with new data sources as well as desperate for better diagnostics, the application of spatial and visual approaches to optical tracking data represents a vital new corridor to new kinds of basketball expertise. Furthermore, perhaps no aspect of basketball is as important and as under-studied as defense. Our case studies were designed to show how new data and emerging approaches can be integrated to help analysts better characterize defense in the NBA. While we contend it is clear that these studies effectively demonstrated the potential of spatial/visual analytics to expose new insights about defense, we also assert that the paper's methods only represent a small first step in a multi-step progression towards the core objective of better defensive analytics.

Evaluating defense in the NBA is very difficult. Despite the new analytical opportunities introduced by player tracking data, our current ability to extract meaningful defensive analytics from these data remains undeveloped. This fundamental notion manifests in multiple ways within our evaluation of interior defense. Perhaps the biggest limitation in our study involves the sample; player-tracking data is only being collected in a subset of NBA arenas. More specifically, as of January 2013, only 15 NBA arenas are equipped with SportVu systems. This obviously biases the sample and is likely to introduce error into our results. But our goal was not to generate the "be-all end-all" ensemble of defensive analytics; instead our goal was to demonstrate the viability of spatial approaches as they relate to making sense of defensive performance data.

Another key limitation is the lack of context associated with the data. Optical tracking data enables us to track player movements in fascinating new ways, but it also reduces players to geometric primitives that frequently obscure the nature of an action. In reality we know players are not coordinate pairs, they are athletic human beings. When we reduce Serge Ibaka to a simple x,y pair, we lose key information. In reality, Serge Ibaka is a 3-dimensional creature with arms that stretch and legs that jump. While this is painfully obvious, even our most sophisticated player tracking systems model NBA players as discrete locations on a plane. This dramatic abstraction of reality introduces infinite issues relating to uncertainty and error. Although we contend there is a vast amount of value in optical tracking data, more research is needed to evaluate uncertainty and reliability in these kinds of investigations.

Conclusion

This paper has sought two accomplish two main objectives: 1) demonstrate that the combination of spatial analyses, visual analytics, and optical tracking data presents a potent new mechanism to understand defensive effectiveness in the NBA, and 2) expose important challenges associated with measuring defensive performances in the NBA. Despite some relevant limitations, we contend that our results suggest that interior defensive abilities vary considerably across the league; simply stated, some players are more effective interior defenders than others. In terms of affecting shooting, we evaluated interior defense in 2 separate case studies. Each study focused on important aspects of interior defense, and as a result each study both answers and provokes important questions about defensive analytics. Although we acknowledge that neither study clearly identifies the best and worst interior defenders, we also contend that 1) each study effectively reveals important characteristics of good defensive play, and 2) advancing defensive analytics will be an long-term iterative process that will require several investigations and multiple new approaches. Lastly, due to his outstanding performance in both case studies, we conclude by suggesting Larry Sanders is the best interior defender in the NBA.

Acknowledgements

The authors would like to acknowledge the help and support of Brian Kopp, Ryan Warkins, Ryan Shea, and David Sherman of STATS. Thank you!

Appendix 1: Expanded Results from Case Study 1: Basket Proximity Shots faced when defender was within 5 feet of basket.

Rank	Defender	Shots Faced	% Close Range	% Mid-range	% 3-point range	**Close FG%**	Mid FG%	3-point FG%
1	Roy Hibbert	419	54.4	29.6	14.8	38.2	37.9	30.7
2	**Larry Sanders**	**622**	**61.9**	**22.2**	**15.4**	**38.4**	**32.6**	**30.2**
3	Elton Brand	198	57.1	26.8	14.1	39.8	32.1	46.4
4	**Serge Ibaka**	**104**	**74.0**	**16.3**	**9.6**	**41.6**	**35.3**	**10.0**
5	LaMarcus Aldridge	221	58.8	24.9	14.5	43.9	38.2	46.9
6	Jermaine O'Neal	392	56.9	28.1	14.0	44.0	32.7	32.7
7	Kosta Koufos	200	60.0	23.5	15.5	45.0	31.9	25.8
8	Kendrick Perkins	745	59.3	24.3	16.1	45.5	37.0	36.7
9	Joakim Noah	334	56.6	27.8	14.4	45.5	44.1	31.3
10	**Dwight Howard**	**409**	**48.2**	**32.0**	**19.1**	**45.7**	**38.2**	**43.6**
11	JaVale McGee	401	53.6	30.2	16.0	46.1	40.5	40.6
12	Amir Johnson	207	56.5	25.1	17.4	46.2	44.2	41.7
13	Ekpe Udoh	468	65.2	20.1	14.1	46.2	42.6	34.9
14	Andris Biedrins	317	49.8	28.7	20.5	46.8	42.9	36.9
15	Tim Duncan	930	57.3	28.6	13.7	47.1	41.4	46.5
16	Emeka Okafor	310	52.3	24.5	22.6	47.5	43.4	32.9
17	Jeremy Tyler	177	60.5	29.4	10.2	47.7	38.5	50.0
18	Nick Collison	273	52.0	27.8	18.7	47.9	35.5	37.3
19	Kevin Seraphin	475	55.6	28.4	14.9	48.1	40.7	35.2
20	DeMarcus Cousins	279	48.4	29.7	20.1	48.2	38.6	44.6
21	Marcus Camby	204	57.8	28.4	13.7	48.3	41.4	39.3
22	Kevin Garnett	772	54.4	28.6	16.2	48.3	37.6	41.6
23	Tiago Splitter	687	58.8	27.2	13.5	48.5	38.0	35.5
24	Samuel Dalembert	777	56.4	30.4	12.9	48.6	38.6	41.0
25	Nene Hilario	212	56.1	25.9	17.9	48.7	36.4	29.0
26	Aaron Gray	275	54.5	29.5	15.6	49.3	39.5	44.2
27	Ed Davis	391	61.6	23.3	14.8	49.4	40.7	34.5
28	Nazr Mohammed	260	51.9	28.1	18.8	49.6	41.1	28.6
29	Chris Bosh	263	55.9	23.2	19.8	49.7	32.8	40.4
30	Marcin Gortat	679	60.2	27.0	12.5	50.4	42.6	37.7
31	Al Jefferson	340	52.9	29.7	16.8	50.6	38.6	33.3
32	Jonas Valanciunas	239	59.4	26.8	13.4	50.7	43.8	37.5
33	Omer Asik	578	58.8	25.8	14.0	51.2	39.6	29.6
34	Greg Stiemsma	407	56.0	31.2	12.5	51.3	41.7	33.3
35	Tyson Chandler	794	57.6	25.6	16.5	51.4	40.4	34.4
36	Nikola Vucevic	421	64.4	24.0	11.6	51.7	37.6	26.5
37	Marc Gasol	344	52.6	26.5	19.5	51.9	33.0	32.8
38	Spencer Hawes	185	61.1	29.7	8.6	52.2	32.7	25.0
39	Nikola Pekovic	669	55.5	27.5	16.9	52.6	46.2	38.1
40	Greg Smith	207	65.2	26.1	8.2	52.6	35.2	35.3
41	Tristan Thompson	174	67.2	18.4	14.4	53.0	56.3	32.0
42	Chris Wilcox	217	65.9	21.2	12.4	53.2	45.7	40.7
43	Robin Lopez	201	61.2	22.4	14.9	53.7	31.1	33.3
44	Jordan Hill	195	60.5	17.9	20.5	54.2	51.4	27.5
45	Tyler Zeller	298	53.4	30.5	16.1	54.7	40.7	33.3
46	Chris Kaman	445	51.2	25.6	22.5	54.8	42.1	42.0
47	Drew Gooden	562	61.6	23.3	14.9	54.9	37.4	36.9
48	Anderson Varejao	224	56.7	23.2	19.2	55.9	42.3	37.2
49	Kevin Love	357	55.2	24.6	20.2	57.9	36.4	31.9
50	Greg Monroe	302	57.0	25.2	17.5	58.7	55.3	35.9
51	David Lee	400	60.3	23.3	15.3	61.0	33.3	29.5
52	Luis Scola	199	62.8	22.6	14.6	62.4	28.9	27.6

Appendix 1A: Same data as Appendix 1, but sorted according to % of shots occurring close to the basket

Rank	Defender	Shots Faced	**% Close Range**	% Mid-range	% 3-point range	Close FG%	Mid FG%	3-point FG%
1	**Dwight Howard**	**409**	**48.2**	**32**	**19.1**	**45.7**	**38.2**	**43.6**
2	DeMarcus Cousins	279	48.4	29.7	20.1	48.2	38.6	44.6
3	Andris Biedrins	317	49.8	28.7	20.5	46.8	42.9	36.9
4	Chris Kaman	445	51.2	25.6	22.5	54.8	42.1	42
5	Nazr Mohammed	260	51.9	28.1	18.8	49.6	41.1	28.6
6	Nick Collison	273	52	27.8	18.7	47.9	35.5	37.3
7	Emeka Okafor	310	52.3	24.5	22.6	47.5	43.4	32.9
8	Marc Gasol	344	52.6	26.5	19.5	51.9	33	32.8
9	Al Jefferson	340	52.9	29.7	16.8	50.6	38.6	33.3
10	Tyler Zeller	298	53.4	30.5	16.1	54.7	40.7	33.3
11	JaVale McGee	401	53.6	30.2	16	46.1	40.5	40.6
12	**Roy Hibbert**	**419**	**54.4**	**29.6**	**14.8**	**38.2**	**37.9**	**30.7**
13	Kevin Garnett	772	54.4	28.6	16.2	48.3	37.6	41.6
14	Aaron Gray	275	54.5	29.5	15.6	49.3	39.5	44.2
15	Kevin Love	357	55.2	24.6	20.2	57.9	36.4	31.9
16	Nikola Pekovic	669	55.5	27.5	16.9	52.6	46.2	38.1
17	Kevin Seraphin	475	55.6	28.4	14.9	48.1	40.7	35.2
18	Chris Bosh	263	55.9	23.2	19.8	49.7	32.8	40.4
19	Greg Stiemsma	407	56	31.2	12.5	51.3	41.7	33.3
20	Nene Hilario	212	56.1	25.9	17.9	48.7	36.4	29
21	Samuel Dalembert	777	56.4	30.4	12.9	48.6	38.6	41
22	Amir Johnson	207	56.5	25.1	17.4	46.2	44.2	41.7
23	Joakim Noah	334	56.6	27.8	14.4	45.5	44.1	31.3
24	Anderson Varejao	224	56.7	23.2	19.2	55.9	42.3	37.2
25	Jermaine O'Neal	392	56.9	28.1	14	44	32.7	32.7
26	Greg Monroe	302	57	25.2	17.5	58.7	55.3	35.9
27	Elton Brand	198	57.1	26.8	14.1	39.8	32.1	46.4
28	Tim Duncan	930	57.3	28.6	13.7	47.1	41.4	46.5
29	Tyson Chandler	794	57.6	25.6	16.5	51.4	40.4	34.4
30	Marcus Camby	204	57.8	28.4	13.7	48.3	41.4	39.3
31	LaMarcus Aldridge	221	58.8	24.9	14.5	43.9	38.2	46.9
32	Tiago Splitter	687	58.8	27.2	13.5	48.5	38	35.5
33	Omer Asik	578	58.8	25.8	14	51.2	39.6	29.6
34	Kendrick Perkins	745	59.3	24.3	16.1	45.5	37	36.7
35	Jonas Valanciunas	239	59.4	26.8	13.4	50.7	43.8	37.5
36	Kosta Koufos	200	60	23.5	15.5	45	31.9	25.8
37	Marcin Gortat	679	60.2	27	12.5	50.4	42.6	37.7
38	David Lee	400	60.3	23.3	15.3	61	33.3	29.5
39	Jeremy Tyler	177	60.5	29.4	10.2	47.7	38.5	50
40	Jordan Hill	195	60.5	17.9	20.5	54.2	51.4	27.5
41	Spencer Hawes	185	61.1	29.7	8.6	52.2	32.7	25
42	Robin Lopez	201	61.2	22.4	14.9	53.7	31.1	33.3
43	Ed Davis	391	61.6	23.3	14.8	49.4	40.7	34.5
44	Drew Gooden	562	61.6	23.3	14.9	54.9	37.4	36.9
45	**Larry Sanders**	**622**	**61.9**	**22.2**	**15.4**	**38.4**	**32.6**	**30.2**
46	Luis Scola	199	62.8	22.6	14.6	62.4	28.9	27.6
47	Nikola Vucevic	421	64.4	24	11.6	51.7	37.6	26.5
48	Ekpe Udoh	468	65.2	20.1	14.1	46.2	42.6	34.9
49	Greg Smith	207	65.2	26.1	8.2	52.6	35.2	35.3
50	Chris Wilcox	217	65.9	21.2	12.4	53.2	45.7	40.7
51	Tristan Thompson	174	67.2	18.4	14.4	53	56.3	32
52	**Serge Ibaka**	**104**	**74**	**16.3**	**9.6**	**41.6**	**35.3**	**10**

Appendix 2: Expanded Results from Case Study 2: Shot Defended: Shots faced when defender was close to shooter.

A) **Shots defended:** The results are presented as:
(% of shots where defenders was within 1 foot) / (% within 3-feet) / (% within 5-feet)

5-ft Rank	Defender	Shots Faced	Within 1ft	Within 3ft	Within 5ft
1	Josh Harrellson	206	1.5	22.3	35.9
2	Kosta Koufos	447	3.8	21.3	35.6
3	Jordan Hill	519	1.9	22.4	35.1
4	**Serge Ibaka**	**223**	**1.8**	**21.1**	**34.5**
5	Chris Wilcox	616	1.6	17.9	34.4
6	Greg Smith	613	1.6	16.6	34.3
7	Jermaine O'Neal	925	1.6	17.6	34.1
8	Cole Aldrich	364	0.8	14.0	33.8
9	Greg Stiemsma	910	1.9	18.6	33.7
10	Jonas Valanciunas	536	2.1	18.1	33.6
11	Ekpe Udoh	1321	2.0	17.6	32.6
12	**Larry Sanders**	**1482**	**2.4**	**17.3**	**32.5**
13	Spencer Hawes	553	1.4	16.3	32.2
14	Jason Collins	207	1.9	16.9	31.9
15	Jeremy Tyler	502	2.4	18.1	31.7
16	Marcin Gortat	1745	1.7	17.3	31.4
17	Elton Brand	547	1.3	17.0	31.3
18	Gustavo Ayon	437	0.9	16.5	31.1
19	Robin Lopez	540	0.7	16.9	31.1
20	Tim Duncan	2353	1.4	17.5	31.1
21	Kevin Love	922	1.8	15.2	31.0
22	Amir Johnson	488	2.9	17.4	30.5
23	Drew Gooden	1513	1.2	13.4	30.3
24	Tiago Splitter	2022	1.6	15.4	30.1
25	Anthony Randolph	284	0.4	15.5	29.9
26	Andray Blatche	389	1.5	14.4	29.8
27	Jon Leuer	239	1.3	15.5	29.7
28	Lavoy Allen	239	1.7	17.6	29.7
29	Amar'e Stoudemire	253	2.8	13.0	29.6
30	Chris Kaman	1095	2.5	17.4	29.5
31	Kurt Thomas	235	2.1	13.2	29.4
32	Nikola Vucevic	1352	1.7	14.1	29.0
33	Tyson Chandler	2186	1.3	15.7	29.0
34	Kevin Garnett	2067	0.8	13.0	28.7
35	Andrew Bogut	248	1.2	13.3	28.6
36	**Roy Hibbert**	**1094**	**1.9**	**16.5**	**28.6**
37	Jason Smith	221	1.8	15.4	28.5
38	Brandan Wright	435	0.9	13.8	28.5
39	Andris Biedrins	783	1.9	15.2	28.4
40	Al Jefferson	947	1.6	14.3	28.3
41	Ian Mahinmi	532	2.1	16.4	28.2
42	Ed Davis	1171	1.2	14.6	28.2
43	Tyler Zeller	741	1.1	13.5	27.9
44	Samuel Dalembert	1765	1.9	15.2	27.8
45	Anderson Varejao	603	0.7	12.9	27.5
46	Kevin Seraphin	1141	1.5	16.7	27.5
47	Tristan Thompson	611	1.3	14.1	27.5
48	Darko Milicic	420	1.0	13.8	27.4
49	Patrick Patterson	280	1.8	13.9	27.1
50	Nick Collison	854	1.4	12.6	27.0
51	Jared Jeffries	281	1.1	13.5	27.0
52	Brook Lopez	337	2.4	12.8	27.0
53	Omer Asik	1571	1.5	15.1	27.0
54	Joakim Noah	1017	1.3	13.0	26.8
55	Andre Drummond	210	1.4	10.5	26.7
56	Luis Scola	629	1.3	13.8	26.6
57	David Lee	1269	1.2	11.3	26.5
58	Kwame Brown	242	0.8	12.8	26.4
59	Pau Gasol	227	0.0	0.0	26.4
60	Jamaal Magloire	228	2.6	16.2	26.3
61	Jason Thompson	289	1.4	11.1	26.3
62	Chris Bosh	810	0.6	10.5	26.0
63	Kendrick Perkins	2605	1.3	12.6	25.9
64	Ryan Hollins	337	1.5	11.9	25.5
65	Marc Gasol	1071	1.5	13.7	25.5
66	Nazr Mohammed	699	1.3	13.2	25.5
67	Hasheem Thabeet	444	0.9	11.7	25.5
68	Byron Mullens	442	0.5	13.3	25.3
69	Festus Ezeli	487	1.4	12.9	25.3

70	Marcus Camby	674	0.4	11.1	25.2
71	Nene Hilario	598	2.2	11.2	25.1
72	Aaron Gray	836	0.8	11.4	25.0
73	Emeka Okafor	818	1.2	13.3	24.9
74	Greg Monroe	853	2.2	11.5	24.7
75	Meyers Leonard	211	0.9	12.3	24.6
76	Zaza Pachulia	492	1.8	13.6	24.6
77	Ryan Anderson	387	0.8	11.6	24.5
78	DeMarcus Cousins	802	1.5	11.5	24.4
79	DeAndre Jordan	566	0.5	10.6	24.4
80	DeJuan Blair	353	0.8	9.1	24.4
81	JaVale McGee	928	1.6	11.7	24.4
82	Nikola Pekovic	1980	1.0	13.4	24.0
83	**Dwight Howard**	**1071**	**1.2**	**10.0**	**23.6**
84	Al Horford	269	1.9	10.8	23.0
85	LaMarcus Aldridge	732	0.7	10.9	23.0
86	Enes Kanter	318	2.8	12.3	22.6
87	Boris Diaw	285	1.1	7.7	22.5
88	Andrew Bynum	614	0.5	9.8	22.3
89	Blake Griffin	256	0.8	9.0	21.9
90	Andrea Bargnani	727	1.0	9.8	21.9
91	Brandon Bass	271	1.1	11.8	21.8
92	Brendan Haywood	404	1.0	10.4	21.0
93	Tyler Hansbrough	381	1.3	9.2	20.7

B) Proximal FG%: The results summarize the FG% of opponents when each defender was within 5 feet.

Rank	Defender	FG%
1	**Larry Sanders**	**34.9%**
2	Andrea Bargnani	35.2%
3	Kendrick Perkins	37.3%
4	Elton Brand	38.0%
5	**Roy Hibbert**	**38.7%**
6	Kosta Koufos	39.0%
7	Nene Hilario	40.0%
8	Andris Biedrins	41.0%
9	Greg Stiemsma	41.7%
10	Jermaine O'Neal	42.2%
11	JaVale McGee	42.5%
12	Nazr Mohammed	43.3%
13	Ian Mahinmi	43.3%
14	Tim Duncan	43.4%
15	**Dwight Howard**	**43.5%**
16	Marc Gasol	43.6%
17	Kevin Seraphin	43.6%
18	Jeremy Tyler	44.0%
19	LaMarcus Aldridge	44.1%
20	Aaron Gray	44.5%
21	Kevin Garnett	44.9%
22	DeMarcus Cousins	44.9%
23	Marcus Camby	45.3%
24	Ekpe Udoh	45.4%
25	Nick Collison	45.5%
26	Chris Bosh	45.5%
27	Tiago Splitter	45.6%
28	Tyson Chandler	45.7%
29	Samuel Dalembert	45.7%
30	Joakim Noah	45.8%
31	Omer Asik	46.0%
32	Emeka Okafor	46.6%
33	Tyler Zeller	46.9%
34	Chris Wilcox	47.2%
35	Marcin Gortat	47.5%
36	Spencer Hawes	47.8%
37	Nikola Pekovic	48.0%
38	Al Jefferson	48.5%
39	Ed Davis	48.8%
40	Nikola Vucevic	49.0%
41	Greg Smith	49.1%
42	Robin Lopez	49.4%
43	Chris Kaman	49.5%
44	Greg Monroe	50.2%
45	Tristan Thompson	50.6%
46	Luis Scola	51.5%
47	Drew Gooden	51.8%
48	Kevin Love	52.1%
49	Jonas Valanciunas	52.8%
50	David Lee	53.0%
51	Jordan Hill	53.9%
52	Anderson Varejao	54.2%