Overview, Design concepts, and Details (ODD) for the model that drafts agent-based modeling into sports analytics.

This document describes fundamental aspects of an agent-based simulation of a game of basketball following the ODD protocol developed by Grimm et al. (2006). Figure 1 demonstrates the graphical user interface (GUI) of the model. The left-hand side of the GUI has the basketball court, which is split into segments to reflect the probability of a player scoring a basket from each patch and the areas where players can move. Also, the players and backboards can be seen. User-specified inputs are located on the right-hand panel, with the parameters relating to the belief structure of the players. The different graphs and output displays show various game metrics, all of which are comparable to statistics considered important to analyzing the performance for a game of basketball. NetLogo 5.3 3D (Wilensky, 1999) was utilized to implement the model.

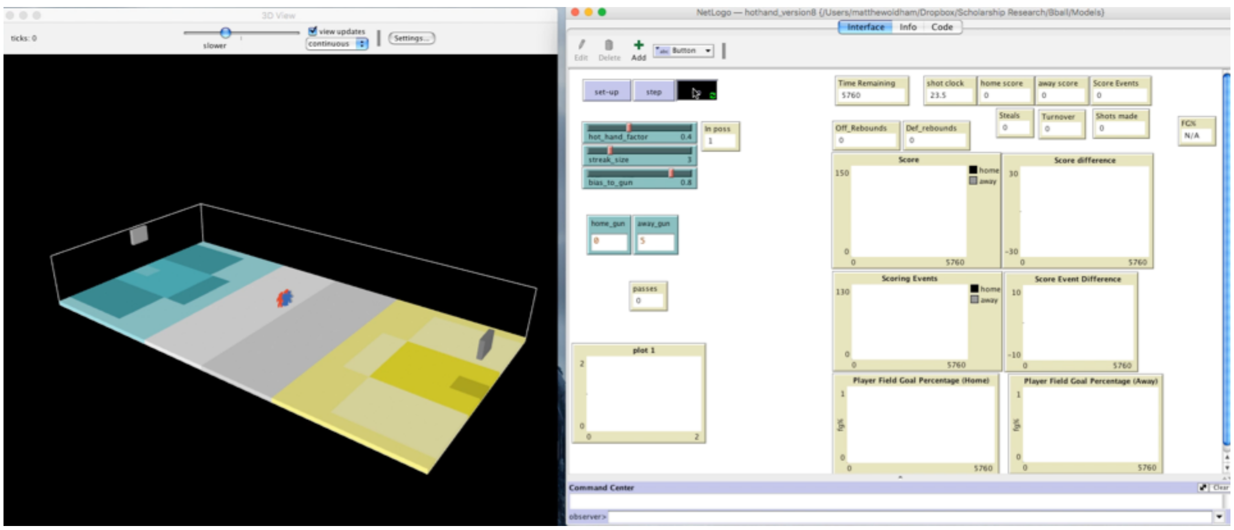

Figure 1: The GUI of the model, which shows the court design and the statistics captured as the model progresses.

1 OVERVIEW

1.1 Purpose

The motivation for implementing an agent-based model (ABM) to simulate basketball comes from the growing literature that proposes the hypothesis that the game dynamics of basketball mimic a complex adaptive system (CAS). The foundation of the approach comes from the fact that an analysis of micro inter-play data shows that the scoring dynamics follows a random walk for a majority of the game, with the exception being the final minute of a quarter, at which time the scoring dynamics produce a power-law distribution (Martín-González, de Saá Guerra, García-Manso, Arriaza, & Valverde-Estévez, 2016). The model's implementation employed the 3D version of NetLogo, with a variety of features used to enhance the realism of the model; such as having the players move around on top of the court and locating hoops at an appropriate height. However, this iteration of the model did not fully exploit the potential of this feature; for example players do not jump off the court.

The model is also designed to agentize the model of Burns (2004), which looked at two specific heuristics that may influence the decisions with regards to the shooting or passing choice of the players involved in a game of basketball. The two heuristics being a team's belief in their franchise (main player) and whether a player has a "hot-hand." The model allows for one player in each team to be a franchise player – think Michael Jordan, that the team recognizes as the "main" player.

The hot-hand debate encapsulates the question of whether a player experiences an extended period of above-average performance, with the term hot-hand referring to the fact that people form (erroneously) the belief that if a player is involved in a successful outcome concerning a random event; for example, making a basket, the probability of future success in that same activity increases. The implication related to sports was that players and supporters formed a belief, based on a limited sample, that successive attempts are positively related. Hence, in basketball if a player has made(missed) 2 shots in a row, the belief is that they are more likely to make(miss) their next shot, that is they have a "hot(cold)-hand" (Gilovich, Vallone, & Tversky, 1985). This research kicked off a 20-year debate regarding the existence of the hot-hand effect. The evidence for the clear majority of sports is that the hot-hand is non-existent, and the random process of the game can partially explain any period of enhanced performance (Bar-Eli, Avugos, & Raab, 2006).

2 State Variables and Scales

Given the intention of the model is to simulate a game of basketball, the key attributes are the court and the players. The court, as detailed in Section 2.2, defines the spatial elements of the model. These elements dictate where the players can move and the probability of making a shot.

The role of the players is to simulate the game within the confines of the court and the various time constraints of the game; that is, the shot clock and the total game time. In making their playing decisions, certain variables are intended to influence the behavior of the players. These variables are:

- The hot hand factor;
- Streak size;
- Bias to gun (belief in the franchise player); and
- Nominated franchise players;

These variables are explained further in Section 2.3, with their direct relevance discussed in Section 5.3. Two other variables that are inputs are the time remaining in the game and the time left on the shot clock. An explanation for the time remaining in the game is straightforward in that it dictates when the simulation is over. The relevance of the shot-clock is that a team must try to score within 24 seconds of gaining possession of the ball, with the clock reset after each shot or turnover. Section 5.3.2 details where and how the shot-clock is significant.

2.1 Global Variables

To allow for the model to proceed as described in Section 3 there are numerous global variables utilized. These variables help determine which procedures are called within each step. Table 1 details the main variables.

Table 1: Details of the global variables.

Variable name	Role
Status	The variable is updated with a value in the interval of -2 to 1 and reflects who has possession of the ball and what action should occur. The values of the status variable and the associated activities are: • 0 – a new play cycle begins after a turnover or a successful shot; • 1 – a shot has been successful; • -1 – a shot just missed, so the rebound procedure is called; and • - 2 – a defensive rebound occurs.
Home_possesion?	This variable tracks whether the home team is in possession of the ball. The role of the variable is to assist in updating the scores and direct the players. For example, when the home team is in possession the players know the location of the home hoop.
Shot?	True or false based on whether the player in possession has decided to shoot. If they have decided to shoot, the shot occurs at the next step in the model.
Pass?	True or false based on whether the player is possession has decided to pass. The pass occurs at the next step in the model.

2.2 The Court

The left-hand side of Figure 1 provides a screenshot of the court design. The dimensions of an National Basketball Association (NBA) basketball court were utilized to inform the court design. The court consists of 98 patches in the y-plain and 50 in the x-plain; therefore, a patch measure a foot in both the x-and-y dimensions. It should be noted that an actual NBA court is 94 feet in length, so the extra 2 feet at each end represent the inbounding area and not additional playing space.

The primary feature of the court design is the varying colors. These colors are important because they define whether the player scores 2 or 3 points, and more importantly, they provide the probability of a player scoring from each position on the court. Section 5.3.2 describes how the players access their court location and use that information in deciding whether to shoot or pass.

Readily available NBA data informed the probabilities for scoring from each location. For example, Mala (2015) provides the details of generating a scoring map for a specific player. The ability to divide the court into separate segments and allocate a specific scoring probability is a demonstration of the potential of using real-world data to inform an ABM. Table 2 provides a summary of the various zones. The court layout also restricts the movements of the players. Section 2.3 discusses this point in more detail.

Table 2: Points scoring per the court design.

Zone	Home team color	Away team colors	Points	Probability of success
Backcourt	N/A	N/A	N/A	0%
Offensive half court (but before the 3-point zone)	White – 2	White – 1	3	2%
3-point zone	Yellow + 2	Cyan + 2	3	33%
Mid-range	Yellow + 3	Cyan - 1	2	43%
Inside the key	Yellow	Cyan	2	50%
The restricted area under the ring	Yellow – 1	Cyan +1	2	70%

2.3 The Players

Primarily, basketball is a game with two teams of 5 players trying to record the highest score possible in the allowable game time. With there being two teams, the team in possession of the ball is considered the offensive side and the team without the ball is the defensive team. The object of the offensive team is to score within 24 seconds, while the defensive team's role is to stop this from occurring. Possession of the ball changes team through one of the following methods a successful basket; a defensive rebound; an intercepted pass; and an offensive foul or a steal. The model considers all these aspects except for the offensive foul, or for that matter any foul. In trying to score the players will move around in their offensive half, with the player in possession tasked with passing; dribbling or shooting the ball.

To allow the play to evolve, each player has numerous variables assigned to them at initiation as well as having access to the amount of time left on the shot clock. Table 2 details the variables that define the characteristics of the players, including their inclination for believing in the hot-hand and their franchise. The user defines the value of franchise players, the belief in the hot-hand, and the length of a streak.

Table 3: The player variables which are set at initiation.

Variable name	Code	Role
ID Number	*ID*	Each player is sequentially assigned an identification number (ID). The home team is allocated the ID number 0 to 4, while 5-9 are for the away team. The numbers are used to help identify the franchise player and determine the allocated playing position of the player.
spot x/y/z		These are the court co-ordinates for a player's offensive position. The coordinates are assigned based on the player's ID number. Once a player's team is in possession, the player heads towards their co-

		ordinate and then moves around this location – as described in Section 5.3.2, until a shot or a turnover.
Home player	*home?*	A flag variable utilized to identify the home side. The home team are players with the ID number 0-4 and have their offensive zone in the "yellow" section.
Opponent		Each player has an opponent allocated to them, and they must guard this player when their team is not in possession of the ball. A player's opponent is assigned as the mirror ID variable; for example, player 0's opponent is player 5.
Hot hand factor	HF_i	Equations 1 and 2 (as detailed in Section 5.3.2) show the utilization of this variable. In summary, the operator utilizes this variable to set the inclination that the players have for considering the hot-hand factor. The range of the variable is 0 (no belief) to 1 (total belief) and is incremented by .01.
Streak size threshold	SS_i	This variable sets the benchmark of the consecutive shots a player must make for them, and their teammates, to conclude that the player is on a hot-streak. Equation 1 highlights the role of this variable in that it helps determine whether the player in possession shoots, dribbles or passes.
Franchise player	F_i	This variable is a binary one, set to 1 for the franchise player and 0 or all others. The variable, via Equation 1 and 2, influences the actions of the player in possession of the ball and the attractiveness of the various passing options.
Belief in thefranchise player	BF_i	The variable sets the inclination players (see Equations 1 and 2) have for favoring/believing in their franchise player. The range of the variable is 0 (no belief) to 1 (total belief) and is incremented by .01.

Given the dynamic nature of basketball, and the abilities of ABMs to allow for near continuous updating, the model updates many of the players' variables at the completion of each step in the simulation. Table 4 provides the definition for these variables.

Table 4: Player variables updated during play.

Variable name	Code	Role
Is the player's team in possession of the ball?	teamposs ?	This variable is updated to reflect whether the agent's team has possession of the ball, which in term determines whether the team is required to play offense or defense.
Is the player in possession of the ball?	poss?	The player in possession of the ball has this variable set to true. If the player is in possession, then they are required to either pass, dribble or shoot. Section 3.3 details how this occurs.

Shooting location	SL_{it}	A variable utilized in deciding the probability of a player taking/making the shot. When choosing whether to shoot the model updates the player's location; that is the players x,y and z coordinates. These details are used to report the color of the patch upon which the player stands. This information is then used to determine the probability of the shot being successful
Shooting probability	SP_{it}	From the color of the patch relevant to the location of the agent in possession of the ball, this variable (the probability of a successful shot) is updated. From Equation 1 this values helps to dictate the action of the player.
Hot streak	HS_{it}	The first part of tracking the performance of each player is this variable. The variable records the number of consecutive shots a player has made. Per Equation 1, it helps determine the action of the player.
Field goal percentage (FG%)	$fg\%_{it}$	The second part of tracking player performance is to update the percentage of successful shots made. It is a popular metric used in sports analytics and helps determine whether the player is a viable passing option.
Team percentage	$team\%_{it}$	This variable is updated to reflect the percentage of points the player has scored for their team. The value is used in Equation 2 to help assess the attractiveness of the player receiving a pass.
Players in proximity		A list that is employed to track teammates that are within a pre-determined cone around the player in possession of the ball. The values are used to assess passing options.
Attractiveness to pass to	$A2P_{it}$	Each offensive player has this variable updated, per Equation 2, to assess whether they are an attractive passing option if the player in possession decides to pass the ball.

3 Process Overview and Scheduling

The model proceeds such that every ½ second (the timeframe of a step) players decide their actions. The timeframe of ½ second was perceived as the optimal timeframe to balance player movements, the time required to assess the game, and the time a shot might take to reach the basket. One team is required to play defense – the team not in possession, while the other team plays offense – the team in possession of the ball. For the defensive unit, their players execute the `defensive` procedure. Figure 2 provides a high-level illustration of how the provides the cycle implemented at each step for the offensive team. A simulation consists of 5760 steps which equates to the length of an NBA game which is 48 minutes. The first step is whether the defensive team steals the ball. Section 5.3 provides the details regarding each procedure.

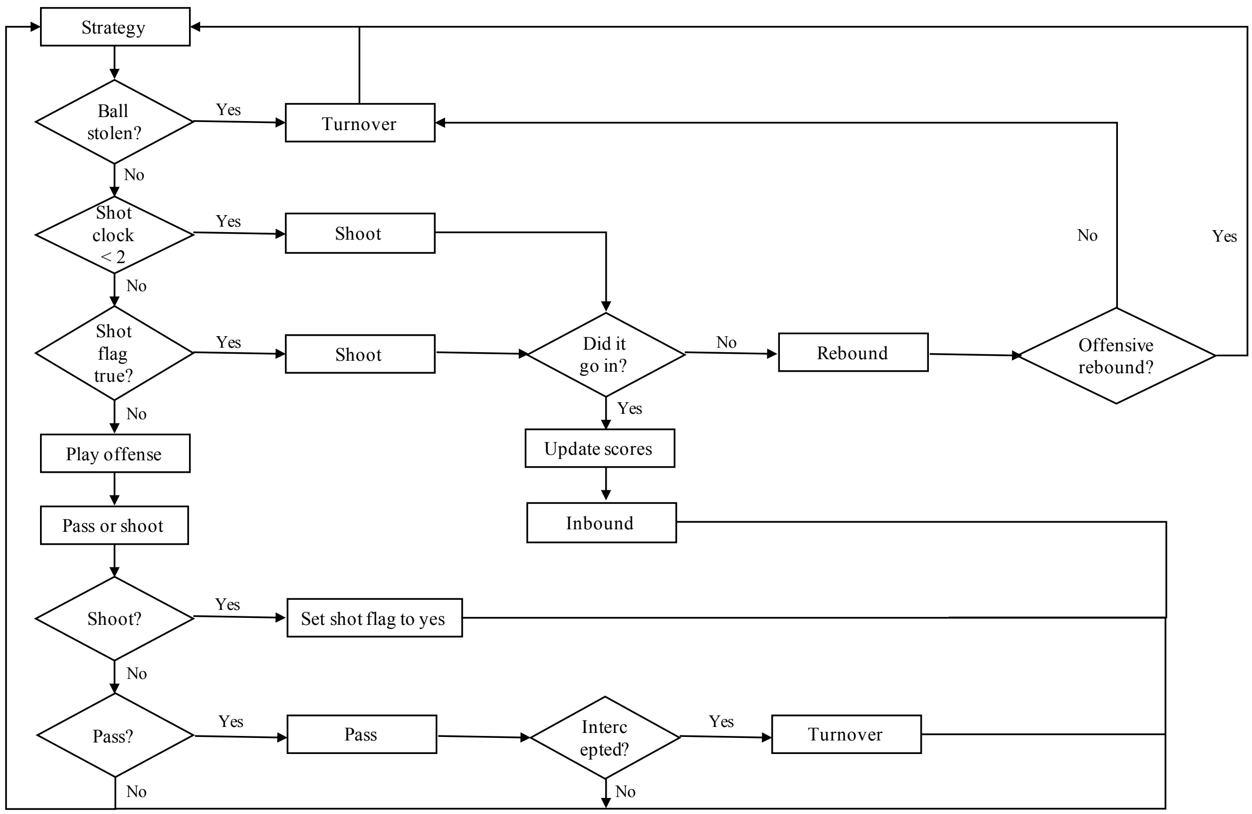

Figure 2: The play cycle.

4 DESIGN CONCEPTS

The ODD protocol calls for various design concepts to be assessed. Table 5 provides the consideration of these concepts in the model.

Table 5: The ODD design concepts and their relevance to the model.

Feature	Utilization
Emergence	Given that players are responsible for completing a “single” play continuously until the time expires, the game score is an emergent outcome. To reinforce this characteristic, it should be noted the players do not consider the game situation in their decision-making process; that is, they do not know whether they are leading or otherwise when deciding their actions.
Adaptation	Adaptation occurs through the players considering the performance of their teammates when deciding whether to shoot, pass or dribble. If the player decides to pass, they also consider performance when deciding the best passing option. Future iterations of the model could look to adjust the decision-making process so that greater adaptation occurs.
Fitness	The players are tasked with scoring before the shot-clock expires. In performing this process a vital component is the player balancing their chance of success with the time remaining on the shot-clock. A player’s decision-making

	considers the past performance of themselves and their teammates. Poor performance will see a player taking fewer shots and not having their teammates pass the ball to them.
Prediction	In predicting whether shooting is the best option, the player considers the probability of the shot attempt being successful. To incorporate this factor the player assesses their location on the court. Other decision cues that are included in predicting the best option are past performance, and the beliefs a team has regarding the hot-hand and their franchise player.
Sensing	As previously stated players can "sense" their location on the court. They are also aware of the performance of themselves and their teammates. An additional sensing component is how the players determine their passing options by scanning a pre-determined zone around themselves.
Interaction	Basketball is a team sport, so the players interact with their teammates. The defensive team interacts with their opponents as they try and block passes.
Stochasticity	As detailed in Section 5.3 multiple sub-models rely on stochastic processes. In summary, these include the probability of making a shot; having a pass intercepted or stolen; and how rebounds the ball.
Collectives	This aspect is covered through the individual players being allocated to one of two teams.
Observation	The model collects a host of game-specific data as the game progresses. This data is then used to verify the model against NBA game data and to test for the effects of the various cognitive biases. The principal metric of concern is the score from each team. Game-data is recorded temporally and at the game-ending level. Other statistics which the model captures include: • Shots taken – with the accompanying percentage made metric calculated, and team scores. A shot taken is recorded as a score event per Figure 3; • Steals; • Rebounds; • Lead changes; and • The maximum lead; The model tracks the above at the team level. At the individual level statistics for each player are also updated, including; • Shots taken – with the accompanying percentage made metric calculated; • Streak size; and • Points scored.

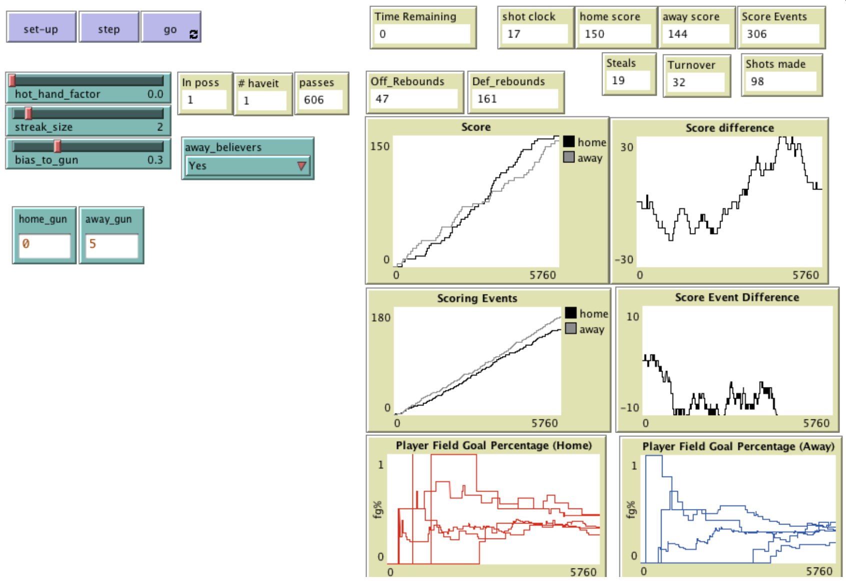

Figure 3: An illustration of the data tracked in real-time by the model.

By way of definitions for Figure 3 – score refers to the actual score of the team, with the score difference being the lead; whereas, a score event records when a team takes a shot, and is therefore a measure the intent of the players rather the result.

5 DETAILS

5.1 Initialization

The first step in the initiation process is to "build" the court. The details of the court are contained in Section 2.2 and remain constant throughout the simulation. Next, the model will create ten players, with five allocated to each of the home and away sides. Per Table 3, the players are assigned various responsibilities at initiation. The general features of the players remain unchanged throughout the simulation. The user can adjust three variables at initiation, these being the belief the players have in the hot-hand effect and their franchise player, plus what the players consider a streak. Again, Table 3 provides a complete description of these variables and their roles. The belief in either cognitive bias, the hot-hand or franchise player, ranges between 0 and 100%. This range is chosen to bound the consideration of the two variables.

5.2 Input Data

The model does not directly rely on any input files. However, the model is informed and calibrated to match real-world data. For instance, as described in Section 2.2 the court is implemented in a manner that replicates an actual NBA. Also, the probability of various events – as described in Section 5.3.2 are informed by actual NBA data.

5.3 Sub-models

The model proceeds such that every ½ second (the timeframe of a step) players decide their actions. This functionality is aided by the values of the `status`, `shot?` and pass? variables.

Depending on the value of these variables the offensive team will dictate the direction of the game. The model identifies the offensive team through the `teamposs?` being setting set to true for all the players in that team. Also, the player in possession of the ball is identified through the `poss?` flag being set to true for that player.

5.3.1 Defense

For the defensive team, their players execute the `defensive` procedure. This procedure works such that as the offensive players move around their allocated position, the defensive players locate their opponent and then set their heading, and move to a location on the defensive side of their opponent; that is, a position between the basket and their opponent. After a shot is taken the defensive team, through the `rebound` procedure, turn and move towards their defensive backboard. The `rebound` procedure, as explained later, determines which team secures the rebound.

5.3.2 Offense

The first step of the model is to call the `strategy` procedure, which takes ½ a second off the shot clock and determines the appropriate action for the offensive team. The next step is whether the defensive team steals the ball. This step sees a random-float drawn between 0 and 1. If the number is less than .03, then the `turnover` procedure is called, and the possession of the ball is switched to the other team and play resets. The value of .03 was chosen to calibrate this metric to the equivalent level of steals from 17 seasons of NBA game data between 2000 and 2016.

If a steal does not occur the status variable is re(set) and the player in possession of the ball will check the shot clock, and if there are less than 2 seconds left they will call the `shoot` procedure. If there is more time left but the player decided to shoot on the previous step, through the `pass_or_shoot` procedure, they will shoot the ball. The `shoot` procedure has players face the hoop of the team in possession, in anticipation of the rebound. For the player taking the shot, they determine their location on the court which is used to set the probability of the shot going in (SP_{it}). To determine whether the shot is successful a random-float between 0 and 1 is compared to the pre-determined probability of the player successfully making the shot from their position on the court.

If the shot is successful; that is, the random-float was less than the probability of the shot going in, the score is updated (reflecting either a 2 or 3-point shot), along with a host of game statistics, including the hot streak tally (HS_{it}), and shots made percentage ($fg\%_{it}$) – which are all vital to determine the future actions of the players. Next, the model calls the `inbound` procedure, which has teams switching possession. At this point, the play process resets, with a new shot clock set and the players move to the other end of the court.

If the shot misses, then the `rebound` procedure is called. Here a random-float between 0 and 1 is called, and if the figure is less than 0.24, the current offensive team rebounds the ball. The ratio of offensive rebounds to defensive rebounds comes from NBA game data. After deciding which team wins the rebound, a player from the appropriate side is randomly chosen to grab the rebound. In the event of an offensive rebound, the play cycle resets, and the shot clock resets. If the defensive team rebounds, the teams switch responsibilities and the shot clock resets. Next, the defensive rebounder passes to their point guard and the team will move to the other end of the court, with the defensive team in pursuit.

On the condition that a player has not shot the ball in the previous step, the team in possession of the ball will enact the `offense` procedure. This procedure moves the offensive players around in their offensive half. To move each player, they add a random value to their x and y co-ordinates of their designated offensive positions. The players then move in the direction of their newly determined offensive position.

The next step is for the player in possession of the ball to decide whether they will shoot, pass or dribble, noting dribbling is the default action if the player does not decide to pass or shoot. In deciding whether to shoot or not, the player in possession establishes where they are on the court. To make this decision the player checks the color of the patch they are currently standing on. From the color of the patch, the player gains an insight into the probability of them making the shot as per the court design described in Section 2.2. Based on the location the player's SP_{it} value is updated. Next, the player checks the time left on the shot clock to determine their desire to shoot based on Equation 1.

$$SorP_{it} = \log((1/TL_{it} * .2) + (SP_{it} * 2))^{(1+ (HS_{it}/SS_i) * HF_i)+(BF_i * F_i))} \quad (1)$$

By way of definitions, Table 1 provides the definition for the variables except for, TL_{it}, which is the time left on the shot clock and SP_{it}, which is the probability of the player making the shot from their current position. The equation is an upward sloping convex function bounded approximately between 0 and 2. The intention being that without the influence of the hot-hand or franchise player there would be an increased probability of the player shooting as the shot clock wound down (that is, TL_{it} decreases) and the higher probability of making the shot (given by SP_{it}).

The effect of the hot-hand (HF_i) and the extra confidence from being the franchise player (F_i) are compounding factors that increase the probability of the player taking the shot. Noting that the F_i variable is binary (1 for the franchise player or 0) so the BF_i variable is only relevant for the franchises player's decision. The influence of the hot-hand effect manifests itself in two ways. The first is the weight that players give the effect through the HF_i (hot-hand factor) variable. The second is through the length of the streak the current player (given through HS_{it}) is on and the threshold that players consider as a streak (SS_i). Therefore, as a player's streak (HS_{it}) grows, the influence of the hot-hand effect grows, with a lower streak threshold amplifying the effect.

After calculating the $SorP_{it}$ value, the player compares it to their shooting threshold. The threshold is drawn from a random-float between 0 and (2 / (1 / TL_{it}). The justification for the approach is that the threshold is a decreasing function concerning the time left on the shot clock; thus, providing a heuristic that a player will have a lower threshold the less time left on the shot clock. The threshold was tuned to calibrate the model, so the number field goal attempts was consistent with a typical NBA game. If the $SorP_{it}$ value is greater than the threshold, the player will set their `shoot?` flag to true and they will take the shot at the next step. If the player decides not to shoot they will continue to dribble or pass, as described next.

If the player has decided not to take the shot, they will call the `pass` procedure. In this procedure, the player in possession will assess the position of their teammates, and decide whether they meet the criteria for being a legitimate passing option. Having established the legitimate passing options, the player will pass to the player with the highest attractiveness. The attractiveness of a player as a passing option is given by Equation 2, with this value updated at the start of the play cycle.

$$A2P_{it} = fg\%_{it} + team\%_{it} + (HS_{it} * HF_i) + (BF_i * F_i) \quad (2)$$

The formula's intention is that the attractiveness of a player as a passing option is dependent on their performance, the team's trust in their franchise player, and their belief in the hot-hand effect. By way of additional definitions, $team\%_{it}$ is the percentage of points the player being assessed has scored for the team; therefore, providing a measure of performance. The influence of the hot-hand effect $(HS_{it} * HF_i)$ and the belief in the franchise player $(BF_i * F_i)$ is linear, with the effect of the franchise player belief being binary. Equation 2 lacks a consideration regarding the defensive pressure the player is experiencing. The lack of concern regarding defensive pressure is also absent in determining the scoring probability.

Before effecting the pass, the model updates the `intercept_options` list. This list includes the opponent of the current dribbler and the player who is about to receive the pass, with one of these players chosen as the possible interceptor. A pass interception occurs if a random-float between 0 and 1 is less than .05 (again calibrated to the average ratio of passes to turnovers in the NBA). A pass interception has the `turnover` procedure called, with teams swapping possession. A successful pass has the receiver status updated to reflect they are in possession of the ball and the play cycle will continue.

6 BIBLIOGRAPHY

Bar-Eli, M., Avugos, S., & Raab, M. (2006). Twenty Years of "Hot Hand" Research: Review and Critique. *Psychology of Sport and Exercise*, *7*(6), 525–553. https://doi.org/10.1016/j.psychsport.2006.03.001

Burns, B. D. (2004). Heuristics as beliefs and as behaviors: The adaptiveness of the "hot hand." *Cognitive Psychology*, *48*(3), 295–331. https://doi.org/10.1016/j.cogpsych.2003.07.003

Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology*, *17*(3), 295–314. https://doi.org/10.1016/0010-0285(85)90010-6

Mala, E. (2015). How to create NBA shot charts in R. Retrieved from https://thedatagame.com.au/2015/09/27/how-to-create-nba-shot-charts-in-r/

Martín-González, J. M., de Saá Guerra, Y., García-Manso, J. M., Arriaza, E., & Valverde-Estévez, T. (2016). The Poisson Model Limits in NBA Basketball: Complexity in Team Sports. *Physica A: Statistical Mechanics and Its Applications*, *464*, 182–190. https://doi.org/10.1016/j.physa.2016.07.028