The Problem of Shot Selection in Basketball

Brian Skinner*

Fine Theoretical Physics Institute, University of Minnesota, Minneapolis, Minnesota, United States of America

Abstract

In basketball, every time the offense produces a shot opportunity the player with the ball must decide whether the shot is worth taking. In this article, I explore the question of when a team should shoot and when they should pass up the shot by considering a simple theoretical model of the shot selection process, in which the quality of shot opportunities generated by the offense is assumed to fall randomly within a uniform distribution. Within this model I derive an answer to the question "how likely must the shot be to go in before the player should take it?" and I show that this lower cutoff for shot quality f depends crucially on the number n of shot opportunities remaining (say, before the shot clock expires), with larger n demanding that only higher-quality shots should be taken. The function $f(n)$ is also derived in the presence of a finite turnover rate and used to predict the shooting rate of an optimal-shooting team as a function of time. The theoretical prediction for the optimal shooting rate is compared to data from the National Basketball Association (NBA). The comparison highlights some limitations of the theoretical model, while also suggesting that NBA teams may be overly reluctant to shoot the ball early in the shot clock.

* E-mail: bskinner@physics.umn.edu

Introduction

In the game of basketball, the purpose of an offensive set is to generate a high-quality shot opportunity. Thus, a successful play ends with some player from the offensive team being given the opportunity to take a reasonably high-percentage shot. At this final moment of the play, the player with the ball must make a decision: should that player take the shot, or should s/he retain possession of the ball and wait for the team to arrive at a higher-percentage opportunity later on in the possession?

The answer to this question depends crucially on three factors: (i) the (perceived) probability that the shot will go in, (ii) the distribution of shot quality that the offense is likely to generate in the future, and (iii) the number of shot opportunities that the offense will have before it is forced to surrender the ball to the opposing team (say, because of an expired shot clock). In this article I examine the simplest analytical model that accounts for all three of these factors and use it to derive a rule for shot selection in basketball.

Recent years have seen something of a revolution in analytical methods in sports, with advanced ideas from game theory, network theory, and statistical mechanics being used to highlight interesting phenomena associated with individual or cooperative performance [1–11]. The problem of shot selection in basketball has been a particularly popular subject of study [6,7,10–15]. Thus far, however, studies have generally focused on either the possible existence of "hot hand" phenomena [7,10,12–14] or on the choice between taking 2- and 3-point shots [6,7,15], and a general theoretical description of the shot-selection process has not been formulated. While the complex nature of decision-making in basketball makes such a description seem prohibitively difficult, it is nonetheless natural to describe the problem of shot selection in basketball as belonging to the class of "optimal stopping problems"

(including, for example, the well-known "secretary problem"), which are often the domain of finance and, more broadly, decision theory and game theory [16].

A very recent work [11] has examined the shot selection problem using the perspective of "dynamic" and "allocative" efficiency criteria. The former criterion requires that every shot be taken only when its quality exceeds the expected point value of the remainder of the possession. The second criteria stipulates that, at optimum, all players on a team should have equal offensive efficiency. This allocative efficiency criterion is a source of some debate, as a recent paper [5] has suggested that the players' declining efficiency with increased usage implies an optimal shooting strategy that can violate the allocative efficiency criterion. Nonetheless, Ref. [11] demonstrates that players in the National Basketball Association (NBA) are excellent at shooting in a way that satisfies dynamic efficiency. That is, players' shooting *rates* seem to be consistent with their shooting *accuracy* when viewed from the requirement of maximizing dynamic efficiency. Still, there is no general theoretical model for addressing the question "when should a shot be taken and when should it be passed up?".

Inspired by these recent discussions, in this article I construct a simple model of the "shoot or pass up the shot" decision and solve for the optimal probability of shooting at each shot opportunity. Within this model, each shot opportunity is characterized by its quality q, which is best defined as the expected number of points that will be scored if the shot as taken; in other words, q is the expected field goal percentage for a given shot multiplied by its potential point value (usually, 2 or 3). If all shots are taken to be worth 1 point, for example, then $0 \leq f_1 \leq f_2 \leq 1$. The possibility of offensive rebounds – whereby the team retains possession of the ball after a missed shot – is not considered explicitly in this article, but one can think that this possibility is lumped into the expected value of a given shot.

Given this definition, I make two important assumptions about the shot quality. The first assumption is that q is a random variable, independent of all other shot opportunities, and is therefore described by some time-independent probability distribution. While this assumption remains somewhat controversial, thus far scoring trends have been shown to be predominately consistent with the assumption of statistical independence between successive shots [7,10,12,17], with a weak "hot hand" effect having been seen only between successive free throw attempts [13,14]. The second major assumption of the model, following Ref. [11], is that the probability distribution for q is a flat distribution: that is, at each shot opportunity q is chosen randomly between some minimum shot quality f_1 and some maximum f_2. This assumption is somewhat arbitrary, and is chosen primarily for the sake of clarity and mathematical simplicity. In principle, however, one can generalize all results presented in this article to the case of a different statistical distribution for q. Some discussion about generalizations and limitations of the model is given at the end of this article in the Discussion section.

The primary concern of this article is calculating a rule for optimizing the shot selection process. That is, this article seeks to derive the optimal minimal value f of the shot quality such that if players shoot if and only if the quality q of the current shot satisfies $q > f$, then their team's expected score per possession will be maximized. It should be noted that this goal of maximizing points per possession is appropriate for "risk neutral" situations, where teams are relatively evenly-matched and a significant amount of time remains in the game. The optimum strategy for end-game or "underdog" situations, where the team tries to maximize (or minimize) the probability of an unlikely upset, is considered in Ref. [15].

One can first note that this "lower cutoff" for shot quality f must depend on the number of plays n that are remaining in the possession. For example, imagine that a team is running their offense without a shot clock, so that they can reset their offense as many times as they want (imagine further, for the time being, that there is no chance of the team turning the ball over). In this case the team can afford to be extremely selective about which shots they take. That is, their expected score per possession is optimized if they hold on to the ball until an opportunity presents itself for a shot that is essentially certain to go in. On the other hand, if a team has time for only one or two shot opportunities in a possession, then there is a decent chance that the team will be forced into taking a relatively low-percentage shot.

So, intuitively, $f(n)$ must increase monotonically with n. In the limit $n = 0$ (when the current opportunity is the last chance for the team to shoot), we must have $f(0) = f_1$; the team should be willing to take even the lowest quality shot. Conversely, in the limit $n \to \infty$ (and, again, in the absence of turnovers), $f(n \to \infty) = f_2$; the team can afford to wait for the "perfect" shot. As I show below, the solution for $f(n)$ at all intermediate values of n constitutes a non-trivial sequence that can only be defined recursively. I call this solution, $f(n)$, "the shooter's sequence"; it is the main result of the present article.

In the following Results section, I present the solution for $f(n)$, use it to derive a relation for the optimal shooting rate as a function of shot clock time, and then compare this theoretical result to data collected from NBA games. The Discussion section uses this comparison to suggest possible suboptimal behaviors among NBA players, and the limitations of the theoretical model are discussed along with some possible generalizations. Finally, the Methods section describes the collection and processing of the NBA data.

Results

1 The shooter's sequence

In this subsection I calculate the optimal lower cutoff for shot quality, $f(n)$, for a situation where there is enough time remaining for exactly n additional shot opportunities after the current one. I also calculate the expected number of points per possession, $F(n)$, that results from following the optimal strategy defined by $f(n)$. The effect of a finite probability of turning the ball over are considered in subsections 2–3.

To begin, we can first consider the case where the team is facing its last possible shot opportunity ($n = 0$). In this situation, the team should be willing to take the shot regardless of how poor it is, which implies $f(0) = f_1$. The expected number of points that results from this shot is the average of f_1 and f_2 (the mean of the shot quality distribution):

$$F(0) = \frac{f_1 + f_2}{2} \qquad (1)$$

Now suppose that the team has enough time to reset their offense one time if they choose to pass up the shot; this is $n = 1$. If the team decides to pass up the shot whenever its quality q is below some value y, then their expected number of points in the possession is

$$F_y(1) = \frac{f_2 - y}{f_2 - f_1} \cdot \frac{y + f_2}{2} + \left(1 - \frac{f_2 - y}{f_2 - f_1}\right) F(0). \qquad (2)$$

In Eq. (2), the expression $(f_2 - y)/(f_2 - f_1)$ corresponds to the probability that the team will take the shot, so that the first term on the right hand side corresponds to the expected points per possession from shooting and the second term corresponds to the expected points per possession from passing up the shot. The optimal value of q, which by definition is equal to $f(1)$, can be found by taking the derivative of $F_y(1)$ and equating it to zero:

$$\left. \frac{dF_y(1)}{dy} \right|_{y = f(1)} = 0. \qquad (3)$$

Combining Eqs. (2) and (3) gives $f(1) = F(0) = (f_1 + f_2)/2$. In other words, the team should shoot the ball whenever the shot opportunity has a higher quality q than the average of what they would get if they held the ball and waited for the next position. This is an intuitive and straightforward result. It can be extended to create a more general version of Eqs. (2) and (3). Namely,

$$F_y(n+1) = \frac{f_2 - y}{f_2 - f_1} \cdot \frac{y + f_2}{2} + \left(1 - \frac{f_2 - y}{f_2 - f_1}\right) F(n). \qquad (4)$$

and

$$\left. \frac{dF_y(n)}{dy} \right|_{y = f(n)} = 0. \qquad (5)$$

Together, these two equations imply

$$f(n+1) = F(n). \tag{6}$$

This is the general statement that a team should shoot the ball only when the quality of the current opportunity is greater than the expected value of retaining the ball and getting n more shot opportunities. In this sense Eq. (6) is quite general, and it is independent of any assumptions about the distribution of available shot opportunities.

The conclusion of Eq. (6) allows one to rewrite Eq. (4) as a recursive sequence for $f(n)$:

$$f(n+1) = \frac{[f(n)]^2 - 2f_1 f(n) + f_2^2}{2(f_2 - f_1)}. \tag{7}$$

Along with the initial value $f(0) = f_1$, Eq. (7) completely defines "the shooter's sequence". Surprisingly, considering the simplicity of the problem statement, this sequence $f(n)$ has no exact analytical solution. Its first few terms and its asymptotic limit are as follows:

$$f(0) = f_1$$

$$f(1) = (f_1 + f_2)/2$$

$$f(2) = (3f_1 + 5f_2)/8$$

$$f(3) = (39f_1 + 89f_2)/128$$

$$f(4) = (8463f_1 + 24305f_2)/32768$$

$$f(n \to \infty) = f_2$$

Note that in the limit where the team has infinite time, their shooting becomes maximally selective (only shots with "perfect" quality f_2 should be taken) and maximally efficient (every possession scores f_2 points).

Since Eq. (7) constitutes a recursive, quadratic map, it has no general solution [18]. Nonetheless, the expression for $f(n)$ can be simplified somewhat by writing it in the form

$$f(n) = \alpha(n)f_1 + \beta(n)f_2, \tag{8}$$

where $\alpha(n)$ and $\beta(n)$ are separate recursive sequences defined by

$$\alpha(n+1) = \alpha(n) - \alpha(n)^2/2, \quad \alpha(0) = 1 \tag{9}$$

and

$$\beta(n+1) = \frac{1 + \beta(n)^2}{2}, \quad \beta(0) = 0, \tag{10}$$

respectively. While $\alpha(n)$ and $\beta(n)$ have no analytical solution, in the limit of large n they have the asymptotic behavior $\alpha(n) \simeq 2/n + \mathcal{O}(1/n^2)$ and $\beta(n) \simeq 1 - 2/n + \mathcal{O}(1/n^2)$.

2 Optimal shooting without a shot clock

In this subsection I consider situations in which there is no natural time limit to a possession, such as informal "pick-up" games. In this case, the number of shot opportunities that the team can generate is limited only by their propensity to turn the ball over – if the team attempts to continually reset the offense in search of a perfect shot they will eventually turn the ball over without taking any shots at all.

Thus, in these situations there is no natural definition of n, which implies that the solution for the optimal shot quality cutoff f is a single number rather than a sequence. Its value depends on the upper and lower values of the distribution, f_1 and f_2, and on the probability p_t that the team will turn the ball over between two subsequent shot opportunities. To calculate f, one can consider that the team's average number of points per possession, F, will be the same at the beginning of every offensive set, regardless of whether they have just chosen to pass up a shot. The team's optimal strategy is to take a shot whenever that shot's quality exceeds F; i.e., $f = F$ as in Eq. (6). This leads to the expression

$$f = p_t \times 0 + (1 - p_t) \left[\frac{f_2 - f}{f_2 - f_1} \cdot \frac{f + f_2}{2} + \left(1 - \frac{f_2 - f}{f_2 - f_1}\right) f \right]. \tag{11}$$

In this equation, the term proportional to p_t represents the expected points scored when the team turns the ball over (zero) and the term proportional to $1 - p_t$ represents the expected points scored when the team does not turn the ball over. As in Eq. (4), the two terms inside the bracket represent the points scored when the shot is taken and when the shot is passed up.

Eq. (11) is a quadratic equation in f, and can therefore be solved directly to give the optimal lower cutoff for shot quality in situations with no shot clock. This process gives

$$f = \frac{f_2 - f_1 p_t - \sqrt{p_t(f_2 - f_1)[2f_2 - p_t(f_1 + f_2)]}}{1 - p_t}. \tag{12}$$

For $0 \leq p_t < 1$ and $0 \leq f_1 \leq f_2$, f is real and positive. In the limit $p_t \to 0$, Eq. (12) gives $f \to f_2$ (perfect efficiency), as expected.

3 The shooter's sequence in the presence of turnovers

In this subsection I reconsider the problem of deriving the shooter's sequence while including the effect of a finite turnover probability p_t. This constitutes a straightforward generalization of Eqs. (4) and (11). Namely,

$$F(n+1) = (1 - p_t) \times$$
$$\left[\frac{f_2 - f(n)}{f_2 - f_1} \cdot \frac{f(n) + f_2}{2} + \left(1 - \frac{f_2 - f(n)}{f_2 - f_1}\right) F(n) \right]. \tag{13}$$

Simplifying this expression and using Eq. (6) gives the recurrence relation

$$f(n+1) = (1-p_t)\frac{f(n)^2 - 2f_1 f(n) + f_2^2}{2(f_2 - f_1)}. \qquad (14)$$

Together with the condition $f(0) = f_1$, Eq. (14) completely defines the sequence $f(n)$.

Unfortunately, the sequence $f(n)$ is unmanageable algebraically at all but very small n. It can easily be evaluated numerically, however, if the values of f_1, f_2, and p_t are known. The first few terms of $f(n)$ and its limiting expression are as follows:

$$f(0) = f_1$$

$$f(1) = (1-p_t)(f_1 + f_2)/2$$

$$f(2) = \frac{1-p_t}{8(f_2-f_1)}\left\{[5-(2-p_t)p_t]f_2^2\right.$$

$$\left. -2f_1 f_2(1-p_t)^2 - f_2^2(1-p_t)(3+p_t)\right\}$$

$$f(n\to\infty) = \frac{f_2 - f_1 p_t - \sqrt{p_t(f_2-f_1)[2f_2 - p_t(f_1+f_2)]}}{1-p_t}$$

Notice that $f(n)$ approaches the result of Eq. (12) in the limit where many shot opportunities remain (*i.e.* the very long shot clock limit).

Overall, the sequence $f(n)$ has two salient features: (i) it increases monotonically with n and ultimately approaches the "no shot clock" limit of Eq. (12), and (ii) it generally calls for the team to accept lower-quality shots than they would in the absence of turnovers, since the team must now factor in the possibility that future attempts will produce turnovers rather than random-quality shot opportunities.

4 Shooting rates of optimal shooters

The preceding subsections give the optimal shot quality cutoff as a function of the number of shots remaining. In this sense, the results presented above are useful for a team trying to answer the question "when should we take a shot?". However, these results do not directly provide a way of answering the question "is the team shooting optimally?". In other words, it is not immediately obvious how the shooter's sequence should manifest itself in shooting patterns during an actual game, where shot opportunities arise continuously in time.

When analyzing the shooting of a team based on collected (play-by-play) data, it is often instructive to look at the team's "shooting rate" $R(t)$. The shooting rate (also sometimes called the "hazard rate" [11]) is defined so that $R(t)dt$ is the probability that a team with the ball at time t will shoot the ball during the infinitesimal interval of time $(t-dt,t)$. Here, t is defined as the time remaining on the shot clock, so that t decreases as the possession goes on. In this subsection I calculate the optimum shooting rate $R(t)$ implied by the results for $f(n)$. This calculation provides a means whereby one can evaluate how much a team's shooting pattern differs from the optimal one.

In order to calculate optimal shooting rate as a function of time, one should assume something about how frequently shot opportunities arise. In this subsection I make the simplest natural assumption, namely that shot opportunities arise randomly with some uniform rate $1/\tau$. For example, $\tau = 4$ seconds would imply that on average a team gets six shot opportunities during a 24-second shot clock. The assumption of a uniform rate of shot opportunities is, in fact, unlikely to describe real data over the entire range of the shot clock, as discussed below in the Discussion section. Nonetheless, it allows one to derive analytically a number of important results. Possible generalizations from this assumption are discussed further at the end of this article in the Discussion section.

I also make the assumption that there is some uniform turnover rate $1/\tau_t$. This assumption can easily be validated by examining turnover rates from NBA games, as discussed below in subsection 5.

Under this set of assumptions, one can immediately write down the probability $P(t,n;\tau)$ that at a given instant t the team will have enough time for exactly n additional shot opportunities. Specifically, $P(t,n;\tau)$ is given by the Poisson distribution:

$$P(n,t;\tau) = \left(\frac{t}{\tau}\right)^n \frac{e^{-t/\tau}}{n!}. \qquad (15)$$

The probability p_t of a turnover between successive shot opportunities is given by

$$p_t = \int_0^\infty \left(1 - e^{-t'/\tau_t}\right) e^{-t'/\tau}\frac{dt'}{\tau} = \frac{\tau}{\tau_t + \tau}. \qquad (16)$$

The integrand in Eq. (16) contains the probability that there is at least one turnover during a time interval t' multiplied by the probability that there are no shot attempts during the time t' multiplied by the probability that a shot attempt arises during $(t',t'+dt')$, and this is integrated over all possible durations t' between subsequent shot attempts. While the upper integration limit in Eq. (16) should in principle be replaced by the total shot clock length t_0, for $\tau \ll t_0$ this limit can safely be set equal to ∞.

In general, for a team deciding at a given time t whether to shoot, the rate of shooting should depend on the proscribed optimal rate for when there are exactly n opportunities left, multiplied by the probability $P(n,t;\tau)$ that there are in fact n opportunities left, and summed over all possible n. More specifically, consider that a team's optimal probability of taking a shot when there are exactly n opportunities remaining is given by $[f_2 - f(n)]/(f_2 - f_1)$, where $f(n)$ is the shooter's sequence defined by Eq. (14). The probability that the team should shoot during the interval $(t-dt,t)$ is therefore given by

$$R(t)dt = \frac{dt}{\tau}\sum_{n=0}^\infty P(n,t;\tau)\frac{f_2 - f(n)}{f_2 - f_1}. \qquad (17)$$

Inserting Eq. (15) gives

$$R(t) = \sum_{n=0}^\infty \frac{t^n e^{-t/\tau}}{\tau^{n+1}n!}\cdot\frac{f_2 - f(n)}{f_2 - f_1}. \qquad (18)$$

Since the sequence $f(n)$ has no analytical solution, there is no general closed-form expression for $R(t)$.

The expected average efficiency (points/possession) of a team following the optimal strategy defined by $R(t)$ can be derived as follows. For a shot taken at time t, the optimal lower cutoff for shot quality, $f(t)$, is given by $f(t) = f_2 - R(t)\tau(f_2 - f_1)$. The corresponding average shot quality $\bar{q}(t) = [f_2 + f(t)]/2$ is given by

$$\bar{q}(t) = f_2 - \frac{f_2 - f_1}{2} R(t)\tau. \qquad (19)$$

To find the expected number of points per possession, one needs to know the probability that a shot will be taken during a given time interval $(t - dt, t)$. This quantity can be written as $S(t; t_0)R(t)dt$, where $S(t; t_0)$ is the probability that the team still has the ball at time t given that it gained possession at time t_0 (the beginning of the shot clock).

$S(t; t_0)$ can be derived by noting that the rate at which the current possession ends, dS/dt, is given by the sum of the shooting rate and the turnover rate multiplied by the probability that the possession has not ended already:

$$\frac{dS}{dt} = S(t; t_0)\left[\frac{1}{\tau_t} + R(t)\right]. \qquad (20)$$

Rearranging this equation and integrating gives

$$S(t; t_0) = \exp\left[-\int_t^{t_0}\left(R(t') + \frac{1}{\tau_t}\right)dt'\right]. \qquad (21)$$

Given this expression for $S(t; t_0)$ one can calculate the expected number of points scored during the possession, F, by integrating the average shot quality at time t multiplied by the probability of a shot being taken during $(t - dt, t)$ over all times t. That is,

$$F = \int_0^{t_0} \bar{q}(t)S(t; t_0)R(t)dt. \qquad (22)$$

While a closed-form analytical expression for F is not possible, Eq. (22) can easily be evaluated numerically.

As an example to illustrate optimal shooting behavior, consider a team that encounters shot opportunities with rate $1/\tau = 1/(4 \text{ seconds})$ and turns the ball over with rate $1/\tau_t = 1/(50 \text{ seconds})$. Using the sequence defined in Eq. (14), one can evaluate numerically the shooting rate implied by Eq. (18). This result is plotted as the black, solid line in Fig. 1a, using $f_2 = 1$ and $f_1 = 0$. In Fig. 1a the optimal shooting rate is plotted as the dimensionless combination $R(t)\tau$, which can be thought of as the probability that a given shot should be taken if the opportunity arises at time t (as opposed to $R(t)$, which is conditional on an opportunity presenting itself). For reference, I also plot the case where there are no turnovers, $\tau_t \to \infty$. One can note that the finite turnover rate causes the optimal shooting rate to increase appreciably early in the shot clock. In other words, when there is a nonzero chance of turning the ball over the team cannot afford to be as selective with their shots.

The rule for optimal shooting can also be expressed in terms of the optimal lower cutoff for shot quality, f, as a function of time.
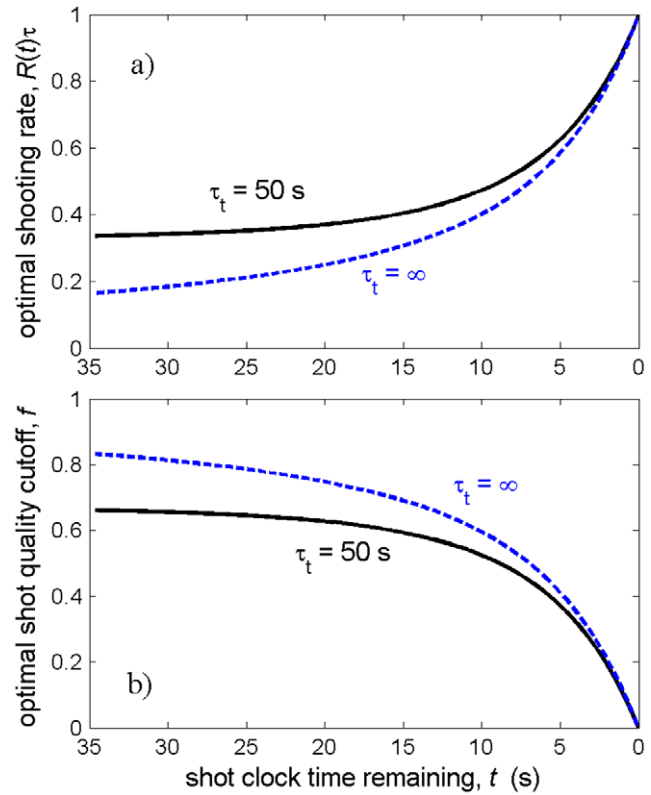
Figure 1. Optimal shooting rate and shot quality cutoff. a) Optimal shooting rate for a hypothetical team with $f_2 = 1$, $f_1 = 0$, $\tau = 4$ seconds, and $\tau_t = 50$ seconds, as given by Eq. (18). The shooting rate $R(t)$ is plotted in the dimensionless form $R(t)\tau$, which can be thought of as the probability that a given shot that has arisen should be taken. The dashed line shows the hypothetical shooting rate for the team in the absence of turnovers. b) Optimal lower cutoff for shot quality, f, as a function of time for the same hypothetical team, both with and without a finite turnover rate.
doi:10.1371/journal.pone.0030776.g001

Since $R(t)\tau$ is the probability that a shot at t should be taken, f can be expressed simply as $f(t) = f_2 - R(t)\tau(f_2 - f_1)$. This optimal lower cutoff is plotted in Fig. 1b. A team that follows the optimal shooting strategy shown in Fig. 1 can be expected to score 0.64 points per possession during games with a 24-second shot clock [see Eq. (22)], a significant enhancement from the value 0.5 that might be naively expected by taking the average of the shot quality distribution.

In the limit of large time t (or when there is no shot clock at all), as considered in subsection 2, the shooting rate $R(t)$ becomes independent of time and Eq. (18) has the following simple form:

$$R = \frac{1}{\tau}\frac{f_2 - f}{f_2 - f_1}, \quad (\text{no shot clock})$$

$$= \frac{1}{\tau_t}\left[\sqrt{1 + \frac{2f_2}{f_2 - f_1}\frac{\tau_t}{\tau}} - 1\right]. \qquad (23)$$

Notice that when turnovers are very rare, $\tau_t \to \infty$, the shooting rate goes to zero, since the team can afford to be extremely selective about their shots.

Eq. (23) also implies an intriguingly weak dependence of the shooting rate on the average time τ between shot opportunities.

Imagine, for example, two teams, A and B, that both turn the ball over every 50 seconds of possession and both have shot distributions characterized by $f_2 = 1$, $f_1 = 0$. Suppose, however, that team A has much faster ball movement, so that team A arrives at a shot opportunity every 4 seconds while team B arrives at a shot opportunity only every 8 seconds. One might expect, then, that in the absence of a shot clock team A should have a shooting rate that is twice as large as that of team B. Eq. (23), however, suggests that this is not the case. Rather, team B should shoot on average every 19 seconds and the twice-faster team A should shoot every 12 seconds. The net result of this optimal strategy, by Eqs. (12) and (16), is that team A scores 0.67 points per possession while team B scores 0.57 points per possession. In other words, team A's twice-faster playing style buys them not a twice-higher shooting rate, but rather an improved ability to be selective about which shots they take, and therefore an improved offensive efficiency.

5 Comparison to NBA data

Given the results of subsection 4, one can examine the in-game shooting statistics of basketball teams and evaluate the extent to which the teams' shooting patterns correspond to the ideal optimum strategy. In this subsection I examine data from NBA games and compare the measured shooting rates and shooting percentages of the league as a whole to the theoretical optimum rates derived above. The data are described in more detail in the Methods section.

The average shooting rate and shot quality (points scored per shot taken) for NBA players are plotted as the symbols in Fig. 2a and b, respectively, as a function of time. Open symbols correspond to shots taken during the first seven seconds of the shot clock, which generally correspond to "fast break" plays during which the offense is not well-described by the theoretical model developed in this article.

In order to compare this data with the theoretical optimum behavior proscribed by the theories of the subsections $1 - 3$, one should determine the values f_1, f_2, τ, and τ_t that best describe the average NBA offense. This last parameter, the average time between turnovers, can be extracted directly from the data: $\tau_t = 100.3$ seconds, as illustrated in Fig. 3. The other parameters can be determined only implicitly, by fitting the observed shooting rates and percentages to the theoretical model.

For the curves shown in Fig. 2, the following approach is employed. First, the average shot quality for NBA teams is determined from the data as a function of time (Fig. 2b). Then, the theoretical average shot quality $[f_2 + f(t)]/2$ of an optimal-shooting team is fit to this data in order to determine the best-fit values of f_1, f_2, and τ, assuming optimal behavior. This procedure gives $f_1 = 0.5$, $f_2 = 1.1$, and $\tau = 2.8$ seconds. The corresponding fit line is shown as the solid curve in Fig. 2b. The shooting rate $R(t)$ implied by these parameter values is then calculated and compared to the shooting rate measured from NBA games (Fig. 2a). In this way one can compare whether the measured shooting *rates* of NBA teams are consistent with their shooting *percentages*, within the assumptions of the theoretical model.

The result, as shown in Fig. 2a, is that NBA teams have a noticeably lower shooting rate during the early stages of the shot clock than is proscribed by the theoretical model. With 15 seconds remaining on the shot clock, for example, the average NBA team has a probability of only about 4% of shooting the ball during the next second, whereas the optimal strategy suggests that this probability should be as high as 12%. This observation is in qualitative agreement with the findings of Ref. [11], which concludes that under-shooting is far more common in the NBA than over-shooting. At small t, the large gap between the observed
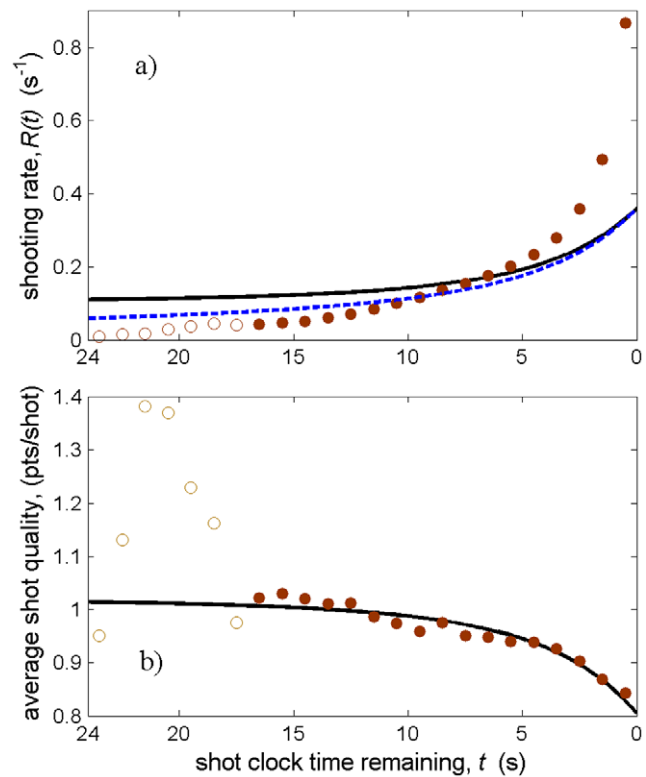

Figure 2. A comparison between the theoretical optimum shooting strategy and data from NBA games. a) The shooting rate as a function of shot clock time t. The solid black line corresponds to the parameters $f_1 = 0.5$, $f_2 = 1.1$, $\tau = 2.8$ s, which are determined by a best fit to the shot quality data, using the NBA average turnover rate $\tau_t = 100.3$ seconds. The dashed blue line corresponds to the same parameters except with the turnover rate $1/\tau_t$ set to zero. b) The average quality (points per shot) of shots taken as a function of the shot clock time t. The solid line corresponds to the best fit curve to the filled symbols, from which the parameters for the solid black line in a) are determined.
doi:10.1371/journal.pone.0030776.g002

and theoretical shooting rates suggests a breakdown of the theoretical model, as discussed in the following section.

The difference between the actual and optimal shooting rates is also reflected in the average scoring efficiency F. For NBA teams, the expected number of points per possession is 0.86, or 0.83 if one considers only possessions lasting past the first seven seconds of the shot clock. In contrast, the optimal shooting strategy shown by the solid lines in Fig. 2 produces 0.91 points/possession for a 24-second shot clock and 0.88 points/possession for a 17-second clock [see Eq. (22)], even though it corresponds to the same distribution of shot quality. This improvement of 0.05 points/possession translates to roughly 4.5 points per game. According to the established "Pythagorean" model of a team's winning percentage in the NBA [19], such an improvement could be expected to produce more than 10 additional wins for a team during an 82-game season.

Discussion

If one operates under the assumption that the theoretical prediction derived in the previous section indeed provides a meaningful comparison with NBA data, then one natural way to interpret the discrepancy between the observed and the theoretically optimal shooting behavior of NBA teams is as a sign of overconfident behavior. That is, NBA players may be unwilling to
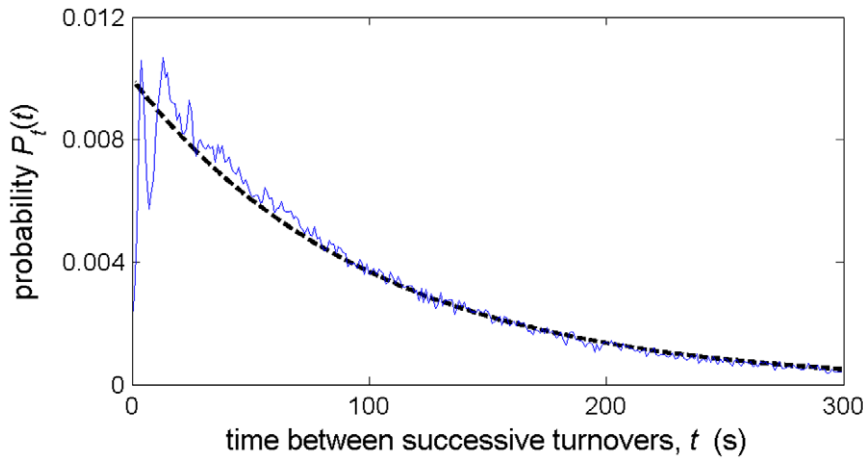
Figure 3. The probability distribution for the time between successive turnovers for NBA teams. The recorded distribution (thin blue line) is well described by the theoretical distribution $P_t(t) = e^{-t/\tau_t}/\tau_t$ (thick black dashed line) corresponding to a uniform turnover rate $1/\tau_t = (100.3 \text{ s})^{-1}$.

settle for only moderately high-quality shot opportunities early in the shot clock, believing that even better opportunities will arise later. Part of the discrepancy can also be explained in terms of undervaluation of turnover rates. If the players believe, for example, that their team has essentially no chance of turning the ball over during the current possession, then they will be more likely to hold the ball and wait for a later opportunity. This effect is illustrated by the dashed blue line in Fig. 2a, which shows the optimal shooting rate for the hypothetical case $\tau_t = \infty$ (the absence of turnovers). This line is in significantly better agreement with the observed shooting rates at large t, which suggests the possibility that when NBA teams make their shooting decisions early in the shot clock they do not account for the probability of future turnovers.

Of course, it is possible that much of the disagreement between the observed and theoretically optimum shooting rates can be attributed to an inaccuracy in the theory's assumption (in subsection 4 above) that shot opportunities arise randomly in time. The breakdown of this assumption can be seen in particular at small t in Fig. 2a, where the shooting rate exceeds $1/\tau$, the supposed rate at which shot opportunities arise. This discrepancy can be seen as an indication that NBA teams often run their offense so as to produce more shot opportunities as the clock winds down, which results in shooting rates that are weighted more heavily toward later times. It is also likely that at very small time t the theory's assumption of a uniform distribution of shot quality becomes invalid. Indeed, in these "buzzer-beating" situations the teams' shots are often forced, and their quality is likely not chosen from the same random distribution as for shots much earlier in the shot clock.

In this sense, the theoretical result of Eq. (18) cannot be considered a very exact description of the shooting rates of NBA teams. In order to improve the applicability of the model for real-game situations, one should account for the possibility of time dependence in the shot quality distribution (f_1 and f_2) and the rate of shot opportunities ($1/\tau$). Such considerations are beyond the scope of the present work.

More generally, a major limitation of the model presented here is its reliance on somewhat arbitrary assumptions about the distribution of shot quality q and about the rate at which these shot opportunities arise. This article has made the simplest set of assumptions – a uniform, time-independent distribution and a uniform rate – but real game situations are likely to be more complex. Unfortunately, the shot quality distribution and rate of generation of shot opportunities cannot be extracted in a straightforward way from existing data. Specifically, game data records only the outcome of shots that were *taken*, and not the quality of opportunities that were passed up. In this way there is no obvious way to gain information about which shot opportunities present themselves without making some assumptions about the players' decision making.

Nonetheless, it should be noted that the model presented here can easily be extended to accommodate different assumptions about the the shot quality distribution and the rate of shot opportunities. Generally speaking, if one assumes that the shot quality distribution is characterized by some probability density function $h(q)$, then the recurrence relation of Eq. (22) becomes

$$f(n+1) = (1 - p_t)\left[\int_{f(n)}^{\infty} qh(q)dq + f(n)\left(1 - \int_{f(n)}^{\infty} h(q)dq\right)\right], \quad (24)$$

so that the entire shooter's sequence can be calculated recursively. [Inserting the flat distribution $h(q) = 1/(f_2 - f_1)$ into Eq. (24) reproduces Eq. (14).] The corresponding optimal shooting rate as a function of time can also be calculated by replacing the Poisson distribution $P(n,t;\tau)$ in Eq. (18) with some other distribution that is assumed to describe the rate at which shot opportunities arise.

Notwithstanding these complications, the model presented in this article nonetheless provides a useful first approach to describing theoretically the problem of shot selection in basketball, and it may be helpful in predicting how optimal strategy should adapt to changing features of the offense – *e.g.* an altered pace of play (τ) or an improving/declining team shooting ability (f_1 and f_2) or a changing turnover rate (τ_t). If nothing else, the theory developed in this article may pave the way for a more complex and accurate theoretical model in the future. In this way the problem of shot selection in basketball should be added to the interesting and growing literature on optimal stopping problems. More broadly, the question of optimal behavior in sports continues to provide an interesting, novel, and highly-applicable playground for mathematics and statistical mechanics.

Methods

The data presented in Figs. 3 and 2 is based on recorded play-by-play data from 4,720 NBA games during the 2006–2007–2009–2010 seasons. These data are available publicly at http://www.basketballgeek.com/data/. From these data, shots taken and points scored are sorted for all possessions based on how much time remains on the shot clock at the time of the shot. Following Ref. [11], possessions that occur within the last 24 seconds of a given quarter or within the last six minutes of a game are eliminated from the data set, since these are less likely to correspond to risk-neutral situations. I also exclude from the data set all shots for which the shot clock time cannot be accurately inferred. These include shots that immediately follow an offensive rebound, defensive foul, or timeout.

The data presented in Figs. 2 and 3 correspond to the average behavior for the NBA as a whole. While a systematic breakdown of shooting rates by team is outside the scope of this article, I note briefly that the shooting rate $R(t)$ is essentially invariant between NBA teams to within statistical noise.

References

1. Oliver D (2004) Basketball on Paper: Rules and Tools for Performance Analysis. Potomac Books, Dulles.
2. Duch J, Waitzman JS, Amaral LAN (2010) Quantifying the performance of individual players in a team activity. PLoS ONE 5: e10937.
3. Sire C, Redner S (2009) Understanding baseball team standings and streaks. The European Physical Journal B - Condensed Matter and Complex Systems 67: 473–481.
4. Radicchi F (2011) Who is the best player ever? a complex network analysis of the history of professional tennis. PLoS ONE 6: e17249.
5. Skinner B (2010) The price of anarchy in basketball. Journal of Quantitative Analysis in Sports 6: Iss.1, Article 3.
6. Alferink LA, Critchfield TS, Hitt JL, Higgins WJ (2009) Generality of the matching law as a descriptor of shot selection in basketball. J Appl Behav Anal 42: 595.
7. Neiman T, Loewenstein Y (2011) Reinforcement learning in professional basketball players. Nature communications 2: Article 569.
8. Raab M, Gula B, Gigerenzer G (2011) The hot hand exists in volleyball and is used for allocation decisions. Journal of Experimental Psychology: Applied.
9. Gabel A, Redner S (2011) Random walk picture of basketball scoring. (unpublished), arXiv: 1109.2825.
10. Wardrop RL (1995) Simpson's paradox and the hot hand in basketball. The American Statistician 49: 24–28.
11. Goldman MR, Rao JM Allocative and dynamic efficiency in nba decision making. MIT Sloan Sports Analytics Conference. Available http://www.sloansportsconference.com.
12. Gilovich T, Vallone R, Tversky A (1985) The hot hand in basketball: On the misperception of random sequences. Cognitive Psychology 17: 295.
13. Arkes J (2010) Revisiting the hot hand theory with free throw data in a multivariate framework. Journal of Quantitative Analysis in Sports 6, Iss.1: Article 3.
14. Yaari G, Eisenmann S (2011) The hot (invisible?) hand: Can time sequence patterns of success/failure in sports be modeled as repeated random independent trials? PLoS ONE 6: e24532.
15. Skinner B (2011) Scoring strategies for the underdog: Using risk as an ally in determining optimal sports strategies. Journal of Quantitative Analysis in Sports 7: Iss.4, Article 11.
16. Moerbeke P (1976) On optimal stopping and free boundary problems. Archive for Rational Mechanics and Analysis 60: 101–148.
17. Huizinga J, Weil S (2009) Hot hand or hot head?the truth about heat checks in the nba. MIT Sloan Sports Analytics Conference.
18. Weisstein EW. Quadratic map. URL http://mathworld.wolfram.com/QuadraticMap.html..
19. Kubatko J, Oliver D, Pelton K, Rosenbaum DT (2007) A starting point for analyzing basketball statistics. Journal of Quantitative Analysis in Sports 3: Iss.3: Article 1.