1 Introduction

Received wisdom among baseball professionals and fans alike holds that it is a clear and obvious advantage to bat in the bottom half of the inning, and therefore have the "privilege" of "last ups." This belief is sufficiently commonplace that no justification is necessary when invoked; for example, after a loss to the Anaheim Angels on August 21, 2002, New York Yankees manager Joe Torre said, "It's tough to lose in an extra-inning game at home because you have the advantage of batting last, but we didn't get the hits when we needed them."[1] Wright and House (1989), in a book taking an analytical approach to the game, assert that batting last is so obvious an advantage that it requires no explicit proof.

The origins of this belief can be traced back to the formative days of the game. Henry Chadwick, the pioneering baseball writer of the nineteenth century, was a strong advocate of the position. In describing a game played by the Brooklyn club at Louisville in 1888 in the (then-Major League) American Association, Chadwick wrote[2]

> The Brooklyn team won their first game of their Western tour today, after a close and exciting contest, in which the rule of being last at the bat was again shown to be of conspicuous advantage... The home team had but one inning left to play, while Brooklyn - owing to being last at the bat - had two, and the confidence the knowledge of this fact gave them was inspiring, and on this occasion, as on others, it gave them the victory.

Note that Brooklyn, despite being the visiting team, batted in the bottom halves of innings in this game. This situation could not typically happen in modern Major League baseball. Since the Official Rules of baseball were codified in their present form in 1950, the visiting team has been designated to bat in the top halves of innings. This is set out in Rule 4.02, which pertains to how a game begins:[3][4]

> The players of the home team shall take their defensive positions, the first batter of the visiting team shall take his position in the batter's box, the umpire shall call "Play" and the game shall start.

Prior to 1950, however, the rules for determining which team batted first varied.[5] Originally, the order in which the teams batted was determined by a coin toss, except in 1877, when the home team batted first. Starting in 1885 in the American Association and in 1887 in the National League, the home team was given the choice of batting first or last. As late as 1894, the home team batted first in 324 of 793 (40.8%) of National League games. Managers of the time were divided on the practice. Philadelphia (managed by Arthur Irwin) batted first 59 times at home; Chicago (Cap Anson) 52 times, and Washington (Gus Schmelz) 49 times. Meanwhile, Cincinnati chose to bat first at home only 6 times, Pittsburgh 4, and Brooklyn only once. By the start of the twentieth century, Irwin, Anson, and Schmelz were no longer managing in the Major Leagues, and the practice of the home team choosing to bat last became standard.

A comment published in the *Detroit News* in 1914 gives one reason why this practice died out:[6]

---

[1] Retrieved from http://www.usatoday.com/sports/scores102/102233/20020821AL–NYYANKEES-0nr.htm/

[2] *Brooklyn Eagle*, June 29, 1888.

[3] The official rules of baseball are available online at the website of Major League Baseball, http://www.mlb.com.

[4] A rare exception occurred in the first game of a doubleheader between the Cleveland Indians and Seattle Mariners in Seattle on September 26, 2007. Seattle's only scheduled visit to Cleveland was wiped out due to cold weather and snow. To compensate, it was decreed that in this game, the Mariners would bat first, despite the game being played in the Mariners' home park. The Mariners were also considered the "home" team for statistical purposes.

[5] See Nemec (1994) for a complete chronology of this rule.

[6] *Detroit News*, September 1, 1914.

Innings	Major Leagues (1960-2006)	Minor Leagues (2006)
10	2547-2359 (.519)	529-501 (.514)
11	1422-1307 (.521)	283-261 (.520)
12	792-730 (.520)	149-146 (.505)
13	454-406 (.528)	81-90 (.474)
14	243-231 (.513)	44-52 (.458)
15	136-127 (.517)	26-30 (.464)
16	72-74 (.493)	13-21 (.382)
17	41-42 (.494)	5-8 (.385)
18	20-24 (.455)	4-4 (.500)
19	12-16 (.429)	3-3 (.500)
20	6-10 (.375)	2-3 (.400)

Table 1: Won-loss record of home teams in games lasting at least N innings.

> Right now all clubs go to the field first when on the home grounds; the custom has become firmly rooted, and no manager ever thinks of changing. Yet, many years ago, it was equally the rule for the home team to bat first, and the argument on which the managers maintained the system was the supposed advantage of 'getting the first crack at the new ball!'
>
> When the game was played with only one ball, and was held up till that ball came back after every journey, a hard-hitting club could, very often, get a flock of runs by starting right in at the jump, taking first bat and collecting hits before the other team had any chance. By the time that ball was turned over to the other club it was black and hard to hit - hence an actual and indisputable advantage for the team first at bat. But when the statute was introduced providing a fresh white ball whenever the original ball vanished, this advantage was destroyed.

It is a robust finding that home teams enjoy an advantage across sports; see Courneya and Carron (1992) for a survey. Because random assignment of teams to the first and last batting roles has rarely been practiced in professional baseball, and not at all in 130 years, natural experiments for separating home field advantage from the effects of the order in which teams bat do not exist in this setting. Other indirect approaches for empirical testing run into similar difficulties. For example, since endgame effects are believed to be central to the last-ups advantage, one would expect that the home team would win a disproportionate number of extra-inning games.

The opposite trend appears in the data. Table 1 presents the won-loss record of home teams in extra-inning games for Major League Baseball since 1960, and Minor League Baseball in 2006.[7] While the number of long games is not large enough to draw statistically significant conclusions, the general trend is for visiting teams as a group to fare better as games get longer. As another datapoint, in a database of over 500 games lasting 20 or more innings drawn from all levels of competition around the world, the visiting team won 56 percent of the time.[8]

Interpreting this as evidence for a strategic advantage to batting first would be premature. The existence of the home field advantage implies that when two teams of equal strength play, the team playing at home will win more than half the time. In Major League Baseball, the overall home team

[7] Data for Major League Baseball is from http://www.retrosheet.org; data for Minor League Baseball is from http://www.minorleaguebaseball.com.

[8] Phil Lowry, personal communication.

winning percentage historically is between .530 and .540. When the game goes to extra innings, this is evidence in favor of the hypothesis that the visiting team is stronger than the home team. This selection bias could easily account for the trend in Table 1.

Because natural experiments for separating home field advantage from the order in which teams bat are lacking in professional baseball, some previous papers seek evidence for an advantage to batting last in baseball and softball games at other levels. Courneya and Carron (1990) studied a municipal softball league in which pairs of teams played two consecutive games on the same day, alternating which team batted last. They found no evidence of an advantage to having the first or last ups. More recently, Simon and Simonoff (2006) sought evidence of an advantage using data from the NCAA baseball and softball tournaments. Many of these games are neutral-site games. In addition, earlier rounds of the tournament are hosted by one of the participating teams; the rules of the tournament result in the host team batting first in some games played at their home site. They find that there is no evidence of an advantage to batting first or last in baseball, and weak evidence that batting first is advantageous in softball.

This paper complements these empirical studies by constructing and solving a model of the progress of a game of baseball as a Markov game. Empirical studies ask whether there is evidence in the field for an advantage to batting first or last. A game-theoretic model permits investigation of the possible mechanisms by which such an advantage might, or might not, manifest itself.

The discrete nature of events in baseball makes Markov modeling a natural idea, which dates at least to Howard (1960). Bellman (1977) observed that baseball strategies could be evaluated by using stochastic dynamic programming methods. The model in this paper extends Bellman's idea by modeling both teams as active strategic actors, making the interaction formally a game instead of a decision problem.

There is a long tradition of quantitative analysis of questions of strategy in baseball, dating to the seminal work of Lindsey (for example, Lindsey (1963)). Most of that study of baseball strategy is situational, taking the continuation values after each possible outcome as given, and asking whether it is advisable, for example, to attempt a sacrifice bunt. For there to be a strategic advantage to batting last (or first), it must be that a team can systematically influence the progress of the game through the choices it makes. Loosely speaking, the existence of an advantage requires that the product of the probability that states arise in which a team has strategic influence, times the magnitude of the effect a team's choices has on the team's chances of ultimately prevailing.

The paper is organized as follows. Section 2 introduces the formal model of baseball as a Markov game. The notion of a "value" of this game is well-defined; all equilibria must result in the same probability of the home team winning, so it is meaningful to ask whether one team or the other has an advantage. Section 3 presents numerical results from solving the model for selected parameterizations calibrated to historical performance in Major League Baseball. Section 4 concludes with discussion and interpretation of the results, and directions for future inquiry.

2 The Model

2.1 A baseline Markov process for baseball

A baseball game is played between two teams, each with nine players. The teams take turns on offense, such that each team has nine turns, or innings, in which it may score runs. After nine innings, the team that has scored the most runs wins the game. If the score is tied after nine innings, an additional extra inning is played, with the team that scores more runs being the winner. If again the game is tied, another extra inning is played, and so on, as necessary, until a winner is determined.

The players on a team take turns batting in a fixed rotation called the batting order. To score a run, a player must advance around a sequence of four bases before the team's turn at bat is terminated. This occurs when three of a team's players have been put "out," which may occur by various means.

The outcome of each player's turn at bat - whether he is put out, or how many bases he advances, or whether he causes another player on a base to be put out, and so forth - is treated as a random event. The progress of a game can then be summarized as a Markov process. In this process, the state is described by a 7-vector $(\iota, \tau, \lambda, \omega, \theta, \beta_1, \beta_2) \in \Sigma$ whose components are

- ι: the current inning (1 through 9; extra innings to break ties are treated as the ninth inning);
- τ: the half inning, denoting whether the team that bats first or second is currently on offense;
- λ: the lead (in runs) currently enjoyed by the batting team (if positive; their deficit in runs if negative);
- ω: the number of outs (0, 1 or 2) recorded so far in the team's inning;
- θ: the set of bases currently occupied by baserunners (a subset of $\{1, 2, 3\}$);
- β_t: the position in the batting order (one through nine) occupied by the next batter scheduled to bat for each team t (the first batter follows the ninth in the rotation).

The transitions between these states occur probabilistically, and are assumed to depend only on β_τ, the identity of the batter, but not any other aspect of the current state. In other words, batters may be heterogeneous, with different batting abilities, but their performance is otherwise independent of the game situation. Because of the structure of the rules of baseball, the number of states which can be reached with positive probability in one transition is small, ranging from 5 to 28.[9]

2.2 Modeling strategic choice

To capture some of the choices available to a baseball team over the course of a game, additional actions are added to the action sets at some states in Σ. Three strategic choices are modeled. The state transitions given are relative to an initial state vector σ, and any components not referenced remain unchanged by the transition.

- The *intentional walk*. At any state σ, the defense may choose to intentionally walk the batter. The subsequent state σ' is given by

$$\lambda' = \begin{cases} \lambda + 1 & \text{if } \theta = \{1,2,3\} \\ \lambda & \text{otherwise} \end{cases}$$

$$\theta' = \begin{cases} \{1\} & \text{if } \theta = \emptyset \\ \{1,2\} & \text{if } \theta = \{1\} \text{ or } \theta = \{2\} \\ \{1,3\} & \text{if } \theta = \{3\} \\ \{1,2,3\} & \text{otherwise} \end{cases}$$

$$\beta'_\tau = \begin{cases} \beta_\tau + 1 & \text{if } \beta_\tau < 9 \\ 1 & \text{if } \beta_\tau = 9 \end{cases}$$

[9] There are five possible transitions from any state with $\theta = \emptyset$: the batter can reach first, second, third, or home, or be put out. The opposite extreme is the case of $\theta = \{1, 2, 3\}$ and $\omega = 0$. Here, there are four players to account for in the succeeding state, as either scoring, being put out, or being still on base after the play.

- The *sacrifice bunt*. If $\theta \in \{\{1\}, \{2\}, \{1,2\}\}$ and $\omega < 2$, the offense may choose to bunt. This play involves the batter deliberately striking the ball softly, so that the runners may advance safely, and the defense has no choice but to put out the batter. The resulting state σ' is given by

$$\begin{aligned} \theta' &= \begin{cases} \{2\} & \text{if } \theta = \{1\} \\ \{3\} & \text{if } \theta = \{2\} \\ \{2,3\} & \text{if } \theta = \{1,2\} \end{cases} \\ \omega' &= \omega + 1 \end{aligned}$$

- The *stolen base*. When $\theta = \{1\}$, a simultaneous-move game is played, which models the stolen base play. This play involves the player on first base attempting to advance to second base (i.e., a state with $\theta = \{2\}$) without waiting for the batter to strike the ball. If successful, the runner has advanced one base closer to his goal of scoring a run; if unsuccessful, the runner costs the team one of its budget of three outs for that inning.

  Table 2 presents the game schematically. The offense chooses between actions S, attempting the play, and N, not attempting the play. The defense chooses between actions B, focusing attention on the batter, and R, focusing attention on the runner who may attempt the play. The effect of action R is an "inspection" action: it reduces the probability of an attempt being successful from one to some probability $0 \leq \rho < 1$. Two sets of results are reported in this paper, one set taking $\rho = .1$, corresponding to relatively ineffective base stealers, and the other taking $\rho = .5$, representing more talented base stealers.

  The subscripted σ represent the continuation state following each contingency of choices. The continuation state σ_S, representing a successful attempt, is characterized by $\theta' = \{2\}$. The continuation state σ_F represents a failed attempt, with $\omega' = \omega + 1$ and $\theta' = \emptyset$. The continuations σ_B and σ_R involve the state transition occurring according to the underlying Markov process. In continuation σ_B, the usual transition probabilities govern the next transition. In continuation σ_R, the transition probabilities shift in favor of the offense; that is, the batter performs better than he would otherwise. This shift operationalizes the idea that the defense pays an inspection "cost." Situational hitting data from Major League Baseballl indicate that the most significant difference in batting with no runners on, versus batting with a runner on first only, is the frequency with which batters hit singles. In this paper, in state σ_R, the probability the batter singles is doubled, and the probability the batter is put out is decreased by the same amount.

  The value of the continuation states are such that the equilibrium of the stage game typically involves randomization by both offense and defense. A companion paper, Turocy (2004), shows that the equilibrium implications of this model are consistent with observed stolen base behavior in Major League Baseball.

2.3 Parameterizations

To approximate the strategic environment found in Major League Baseball, the transition probabilities between any two states σ and σ' are generated using a two-step process, similar to the methods in Trueman (1977) and Bukiet et al (1997). In the first step, a basic outcome for the batter is modeled as a multinomial random variable, with seven possible outcomes: *single*, *double*, *triple*, *home run*, *walk*, *strikeout*, and *generic out*. This last outcome is a catchall for any ball hit in play not resulting

	batter (B)	runner (R)
attempt (S)	σ_S	$\rho\sigma_S + (1-\rho)\sigma_F$
no attempt (N)	σ_B	σ_R

Table 2: A model of the stolen base play as a simultaneous-move game between the offense and the defense. In the table, the offense is the row chooser, and the defense the column chooser. The cell entries are probability distributions over continuation states.

in one of the other outcomes. The frequencies of these outcomes form a vector $\Phi^{ti} = (\phi^{ti}_{1B}, \phi^{ti}_{2B}, \phi^{ti}_{3B}, \phi^{ti}_{\mathrm{HR}}, \phi^{ti}_{\mathrm{BB}}, \phi^{ti}_{\mathrm{SO}}, \phi^{ti}_{\mathrm{OUT}})$ of probabilities representing the batting abilities of batter i on team t. These probabilities are player-specific, capturing the observation that strategic choices in baseball are often said to depend on the characteristics of the players coming up in the batting order,

Once the basic outcome is determined, the state transition itself is determined. Each basic outcome may have several possible state transitions: for example, a player on first base might advance one base on a single, or he might advance two bases. The probabilities of these transitions are set to the overall frequencies in Major League Baseball over the seasons 1973-1992. For example, during the period, there were 111,017 singles hit with no outs ($\omega = 0$) and no runners on base ($\theta = \emptyset$). Of these, 108,434 (97.67%) resulted in a state with $\omega' = 0$ and $\theta' = \{1\}$; 1436 (1.29%) in $\omega' = 0$ and $\theta' = \{2\}$; 261 (0.24%) in $\omega' = 0$ and $\theta' = \{3\}$; 873 (0.79%) in $\omega' = 1$ and $\theta' = \emptyset$; and 13 (0.01%) in $\omega' = 0$, $\theta' = \emptyset$, and $\lambda' = \lambda + 1$. These frequencies then determine the state transition probabilities in the model conditional on a single being hit in that situation.

Although teams differ, there are general patterns in how batting orders are organized in baseball. Stronger hitters tend to bat earlier in the order than weaker hitters, and hitters who hit many extra-base hits tend to bat in the third through fifth positions. Therefore, a "typical" batting order is operationalized by compiling aggregate batting performance by batting order position, using play-by-play data from the 1973 through 1992 seasons of Major League Baseball as published by Retrosheet. The starting point of this period represents the year in which the American League began the use of the designated hitter rule; the pitcher, usually a very weak hitter, rarely bats in the American League, but does bat, almost always in the ninth position, in the National League. Therefore, the vectors Φ are computed separately for the two leagues, since strategy when the bottom of the order is due up is likely to depend on whether the ninth position is occupied by the pitcher.

The paper considers the case of these typical teams in two conditions: games played at a neutral site, and games played at the home site of one of the teams. For games played at a neutral site, the overall frequencies Φ^{ti} are used for both teams, $t = 1, 2$. For games played at the home team's site, the data are further disaggregated into the vectors compiled by batters playing at home, and those compiled by batters playing on the road. The Φ vectors for both conditions for both leagues are included in Appendix A.

2.4 Existence of a well-defined "value"

The focus of this paper is to investigate the case where both teams are making their strategic choices optimally, given the choices made by the opponent. Therefore, the objective is to compute a Nash equilibrium of the game over a suitably-defined strategy space. It is generally possible for a game to have multiple Nash equilibria; in order for the question of a first- or last-ups advantage to have a well-defined answer, it must be that all equilibria of the game have the same "value." That is, all equilibria of the game for a particular parameterization must result in the same winning percentage for the team batting last.

For technical and computational convenience, it will be assumed that there is an upper bound $\Lambda > 0$ on the size of the lead. When the lead of one team reaches Λ, that team is assumed to win with probability one. This can be interpreted as implementing a "mercy rule" similar to those often used to terminate lopsided games in youth and recreational baseball. For the results reported here, $\Lambda = 30$ is chosen. Prior to the Texas Rangers scoring 30 runs against the Baltimore Orioles in August 2007, it had been 110 years since a team in Major League Baseball scored 30 runs in a game; no team has come back to win from a deficit approaching 30.

Additionally, attention will be restricted to Markov, or positional, strategies. In general, a strategy would be a function specifying the choices to make given the entire history of play up to that point. Markov strategies impose the restriction that the strategy depends only on the state $\sigma \in \Sigma$, and not the path of play that resulted in the game reaching that state. Since there are no concepts like "momentum" in this model, this restriction simply rules out strategies depending on payoff-irrelevant factors.

With the mercy-rule assumption, the stochastic game has a finite number of states, and a finite number of actions, and so Markov perfect equilibrium does exist (see, for example, Fudenberg and Tirole (1991), Theorem 13.1). Also, as a finite two-player zero-sum game, all Nash equilibria must give identical payoffs to the players; since Markov perfect equilibria are Nash equilibria, all equilibria of the Markov model of a baseball game must give the same payoff. So, while equilibria may not be unique, it is well-defined to talk about the probability that a team will win the game in (Markov perfect) equilibrium.

Because players bat in a fixed order, and because each player who comes to bat will eventually be put out, score a run, or be on base when the defense records the third out, there is a partial ordering $\succ$ of the states in innings one through eight such that $s_1 \succ s_2$ if and only if state s_1 can occur before state s_2. Therefore, for these innings, it is possible to solve for equilibrium backwards in time, starting with the end of the eighth inning and ending with the beginning of the game.

The process of computing the equilibrium thus spends most of its time solving the fixed-point problem involving the ninth inning, as the value of a tie game at the end of the ninth inning must be equal to the value of a tie game at the beginning of the ninth inning, everything else being equal. Value-function iteration on the ninth inning converged rapidly. Iteration was stopped when the value of a tie game at the beginning of the ninth inning changed by less than 10^{-4} in successive iterations; this threshold was generally reached within 15-20 iterations. Therefore, the value of the game can be computed within a minute, even though the state space consists of 2,134,512 distinct states.

3 Results

Table 3 displays the winning percentage of the team batting last in neutral-site games between identical teams. Table 4 reports the winning percentage of the home team when they bat last, and when they bat first. Each table reports eight cases, in which each of the strategic choices being considered is available or not available, independently of the others. Separate figures are reported for the American and National leagues.

The winning percentages in Tables 3 and 4 imply that for teams of approximately equal strength, which team bats last has only a small effect on the relative likelihoods of eventual victory. For a sense of scale in interpreting winning percentages, note that the schedule of a Major League Baseball team consists of 162 games, 81 of which are played at the team's home park. A difference in winning percentage of .001 corresponds to one extra victory or defeat per six seasons overall. Restricting consideration to only the home portion of a team's schedule, that winning percentage margin of .001 would amount to one extra victory or defeat in twelve seasons of play. The percentages in Tables 3

SB	IBB	SH	American League	National League
no	no	no	.50000	.50000
$\rho = .1$	no	no	.49988	.49988
$\rho = .5$	no	no	.49983	.49982
no	yes	no	.49983	.49972
$\rho = .1$	yes	no	.49970	.49958
$\rho = .5$	yes	no	.49964	.49951
no	no	yes	.50057	.50068
$\rho = .1$	no	yes	.50044	.50046
$\rho = .5$	no	yes	.50038	.50034
no	yes	yes	.50017	.50013
$\rho = .1$	yes	yes	.50003	.49991
$\rho = .5$	yes	yes	.49996	.49976

Table 3: Probability the last-batting team wins a game between identical teams, given the set of stage-game actions available (SB=stolen base, IBB=intentional walk, SH=sacrifice bunt).

			American League		National League	
SB	IBB	SH	last	first	last	first
no	no	no	.53278	.53278	.53547	.53547
$\rho = .1$	no	no	.53180	**.53203**	.53508	**.53532**
$\rho = .5$	no	no	.53138	**.53173**	.53494	**.53530**
no	yes	no	.53260	**.53295**	.53519	**.53575**
$\rho = .1$	yes	no	.53161	**.53222**	.53478	**.53563**
$\rho = .5$	yes	no	.53118	**.53191**	.53464	**.53562**
no	no	yes	**.53325**	.53211	**.53604**	.53468
$\rho = .1$	no	yes	**.53226**	.53138	**.53556**	.53464
$\rho = .5$	no	yes	**.53184**	.53108	**.53537**	.53468
no	yes	yes	**.53284**	.53251	**.53550**	.53524
$\rho = .1$	yes	yes	**.53184**	.53179	.53502	**.53520**
$\rho = .5$	yes	yes	.53141	**.53150**	.53479	**.53528**

Table 4: Winning percentage of home team when batting last or first, respectively, for each of the combinations of stage-game actions available. Cells in bold indicate the higher winning percentage.

and 4 imply that the order in which teams bat, in this model, affects the outcome of a game played by a given team every one or two decades.

The exercise of turning strategies "on" and "off" in Tables 3 and 4, while artificial, helps to illuminate the mechanisms determining the value of the game. Adding the sacrifice bunt always makes batting last more attractive, holding fixed the availability of other strategies. Similarly, adding the intentional walk always makes batting first more attractive. The effect of the stolen base, in which both teams are taking actions, is ambiguous, regardless of whether the offensive players are stronger ($\rho = .5$) or weaker ($\rho = .1$) at stealing bases. These observations are all consistent with a generalized principle: it is advantageous to "go last," in the sense of being the team with the richer set of strategic tools available in the bottom half of an inning. It has traditionally been assumed that the team batting in the bottom half of the inning "goes last." The quantitative results indicate the validity of this assumption is unclear.

The relative effectiveness of strategies depends on the parameterization of Φ. In both the neutral-site and home-site treatments, it is slightly better to bat last in the American League, and first in the National League. While the quantitative differences are not significant in any practical sense, this does illustrate that whether batting first or last is better could depend on the environment and the teams playing.

Finally, these results are driven by endgame effects. The model can be modified to conduct the thought experiment of having games with different numbers of innings than the standard nine. In the American League neutral-site specification, the home team wins with probability .500028 if the game is nine innings; extending this to 20 innings lowers the probability to .500014, and to 100 innings, .500004.

4 Conclusions

Markov perfect equilibrium in a game-theoretic model of baseball suggests that there is no significant advantage to batting first or last. Yet, the belief that such an advantage exists has been widespread for more than a century. There are three possibilities for why this discrepancy exists: (1) the model presented lacks some important strategic characteristic, which, if incorporated, would validate conventional wisdom; (2) conventional wisdom is correct, but for reasons that are non-strategic in nature; (3) the conventional wisdom is wrong, but perceptual biases lead to the propagation of the belief.

The game model analyzed here contains only a small fraction of the possible ways in which baseball teams make choices that affect the outcome of the game. It is possible that omitted factors might change the theoretical outcome. For instance, a significant missing feature of this model is tactical substitution of players. The model presented assumes the nine players on each team cannot be changed. Tactical substitution of players is a salient feature of the later innings of modern professional baseball. It is unclear what the net effect of incorporating substitutions would be. While the team that bats last may be better able to substitute in a better offensive player who is weak defensively in the bottom half of the last inning, at the same time, the team that bats first can substitute in better defensive players if they take a lead into that last half inning. The results from a study of models with tactical substitution would depend on the specific characteristics of the roster of players available to a team. It is likely that the advice such a model would give could depend on these roster characteristics.

While tactical player substitution may be missing from the present analysis, the belief that batting last was an advantage predates the use of such substitutions in baseball. In baseball's original rules, substitutions were not permitted at all. By the time Chadwick wrote the comments quoted in the introduction, limited substitutions were permitted, but most substitutions involved only the replacement of injured players. The significant use of tactical substitutions only began in the second half of the twentieth century. Beliefs about the advantage of batting last were long entrenched prior to this development.

The flip side of this observation is that the last-ups advantage may well have been true in Chadwick's time, but evolution of the game has eroded it. If a modern observer were to be able to watch the first National League game in 1876, he might not recognize its relation to today's sport. For instance, pitchers were required to toss the ball underhanded to the batter, who could request a high or low pitch. Fielders wore no gloves, and the game was played on surfaces of varying quality, far short of today's manicured fields. The calculations in this paper suggest the solution of the last-ups advantage puzzle lies with identifying whether offense or defense has choices which are more effective in affecting the path of play in the game. In early baseball, it may well have been that the offense had at its disposal the preponderance of strategic options.

In modern baseball, much value is placed on the ability of catchers (or, more recently, pitching coaches) to "call" a good game; that is, to select the type and location of pitches in an effective way. Formally modeling this process will be difficult, because of the number of variables involved. Each pitcher has a different repetoire of pitches. Pitches can be targeted to various areas inside or outside the strike zone. It is not clear that a Markov assumption is appropriate in modeling a batter-pitcher confrontation, since some part of effective pitching may involve setting up optical illusions by juxtaposing different types of pitches. Even though a formal model may be impractical, a typical team throws (and faces) 130 to 150 pitches in a game under modern conditions, giving some scope for pitch selection to be a significant strategic variable. If pitch selection is important, the results of this model suggest that a team wants to be on defense in the bottom half of the inning, because the defense would be more in control of the play of the game. It is inconsistent to assign significant value to calling a game while at the same time believing that batting last is a strategic advantage.

This model has proceeded on the assumption that the last-ups advantage is strategic. Alternatively, the advantage may be psychological in nature. Perhaps there is some benefit to the additional certainty of batting when one knows how many runs are needed; or perhaps there is a self-confirming nature to the belief: it's an advantage to bat last because everyone believes there is one. Chadwick used the words 'confidence' and 'inspiring' in his description, suggesting psychology might play a role.

The Courneya and Carron (1990) study of municipal softball leagues found no evidence of an advantage for either team. In slow-pitch softball, the strategies discussed in this paper are likely not relevant: stolen bases and bunting are generally not permitted, the intentional walk is rarely significant, and the use of tactical substitutions is rare. Further, it seems plausible to believe that, if significant psychological effects were to be present, they would be more likely to show themselves in an amateur setting. Players reaching the top level in any professional sport are selected based on performance, which depends on both physical and psychological traits; there is no reason to believe this selection process would favor players prone to having their performance affected by such psychological factors.

It is possible that there are perceptual or framing biases active in sustaining the belief that batting last is an advantage. For instance, close games which are decided at the very end are memorable. When a team scores a run in the bottom half of the last inning to win the game, that fact will appear in the lead paragraph of the game story in the next day's newspaper. A team scoring a run in the top half of the inning, then holding on for the victory, will be less memorable, because the connection between the scoring of the run and the victory is less immediate. A related possibility is that there is an implicit assumption that, because only one team has the ability to score in a half inning, that they must have "control" over how that inning progresses. In practice, a run saved is worth about the same as a run scored, in terms of the overall chances of winning a game. Runs scored are directly observed; runs saved, by definition, are not observed. These gains and losses may be processed in different ways mentally.

In all the baseball and softball applications surveyed by the literature, teams are assigned to first- and last-batting roles. Since teams have no choice in the matter, the question of whether it is an advantage to bat first or last might seem academic. There is a way in which an errant belief in an advantage to batting last might have a negative impact on a team's chances of winning. The quantitative results of the model suggest that a manager would do well in ignoring whether it is the top half or the bottom half of the inning, and basing tactical decisions on the inning, configuration of baserunners, and number of outs. If a manager believed in an advantage to batting last, he might not behave in that way; in fact, a common saying is to "play for the tie at home, and the win on the road." If, in fact, batting last is not an advantage, this would be poor advice. Such a hypothesis is consistent with the pattern of visiting teams tending to win longer games. With the recent development of the Retrosheet play-by-play database, it is now possible to start systematically investigating how strategic decisions depend on game situations in Major League Baseball. A direction for future research is to compare the advice of theoretical models such as this one to the observed behavior in the field.

A Calibrating parameters

i	ϕ_{1B}	ϕ_{2B}	ϕ_{3B}	ϕ_{HR}	ϕ_{BB}	ϕ_{SO}	ϕ_{OUT}
1	.185	.039	.009	.013	.085	.121	.548
2	.175	.042	.007	.017	.082	.117	.560
3	.169	.045	.006	.029	.086	.129	.536
4	.155	.043	.004	.037	.090	.149	.523
5	.157	.043	.005	.031	.085	.148	.531
6	.156	.041	.005	.028	.082	.152	.536
7	.158	.041	.005	.024	.076	.151	.545
8	.164	.037	.005	.017	.073	.145	.558
9	.168	.035	.005	.012	.073	.141	.567

Table 5: Neutral-site batting parameters, American League

i	ϕ_{1B}	ϕ_{2B}	ϕ_{3B}	ϕ_{HR}	ϕ_{BB}	ϕ_{SO}	ϕ_{OUT}
1	.185	.039	.010	.012	.083	.125	.546
2	.180	.040	.007	.012	.076	.114	.571
3	.171	.044	.007	.026	.088	.134	.530
4	.160	.044	.005	.036	.086	.149	.519
5	.162	.043	.005	.029	.079	.143	.539
6	.164	.041	.005	.023	.077	.145	.545
7	.164	.041	.005	.019	.072	.142	.558
8	.166	.035	.006	.010	.068	.140	.575
9	.126	.024	.003	.007	.057	.285	.498

Table 6: Neutral-site batting parameters, National League

Side	i	ϕ_{1B}	ϕ_{2B}	ϕ_{3B}	ϕ_{HR}	ϕ_{BB}	ϕ_{SO}	ϕ_{OUT}
	1	.183	.038	.008	.013	.080	.125	.553
	2	.173	.041	.006	.017	.078	.122	.564
	3	.169	.044	.005	.028	.084	.133	.537
	4	.156	.041	.003	.037	.087	.150	.525
Away	5	.158	.042	.005	.030	.082	.152	.532
	6	.157	.039	.005	.028	.080	.154	.537
	7	.157	.040	.005	.024	.074	.154	.546
	8	.163	.037	.005	.017	.072	.148	.559
	9	.170	.034	.005	.012	.069	.144	.566
	1	.187	.040	.010	.014	.090	.116	.543
	2	.176	.043	.007	.018	.087	.113	.556
	3	.169	.047	.006	.030	.088	.124	.535
	4	.153	.044	.005	.037	.092	.148	.521
Home	5	.156	.044	.006	.033	.089	.144	.529
	6	.156	.043	.005	.028	.085	.150	.534
	7	.159	.042	.006	.024	.079	.148	.543
	8	.166	.038	.005	.017	.075	.143	.557
	9	.166	.035	.006	.011	.076	.137	.568

Table 7: Home site batting parameters, American League

Side	i	ϕ_{1B}	ϕ_{2B}	ϕ_{3B}	ϕ_{HR}	ϕ_{BB}	ϕ_{SO}	ϕ_{OUT}
	1	.181	.038	.009	.012	.080	.131	.548
	2	.176	.040	.006	.012	.072	.117	.576
	3	.172	.044	.006	.025	.086	.138	.529
	4	.161	.043	.004	.035	.084	.152	.521
Away	5	.162	.042	.004	.028	.078	.144	.541
	6	.164	.039	.005	.024	.075	.149	.545
	7	.161	.040	.005	.018	.070	.144	.562
	8	.164	.035	.006	.011	.066	.143	.577
	9	.125	.024	.003	.007	.053	.292	.497
	1	.188	.040	.010	.013	.087	.118	.544
	2	.185	.040	.008	.013	.079	.110	.565
	3	.170	.045	.007	.027	.091	.129	.531
	4	.160	.046	.006	.038	.088	.147	.516
Home	5	.163	.044	.005	.029	.080	.142	.536
	6	.163	.043	.006	.022	.079	.141	.546
	7	.167	.042	.006	.019	.074	.139	.554
	8	.169	.036	.006	.010	.071	.137	.572
	9	.127	.024	.003	.007	.060	.278	.500

Table 8: Home site batting parameters, National League

References

[1] Bellman, R. Dynamic programming and Markovian decision processes, with application to baseball. In S. P. Ladany and R. E. Machol, editors, *Optimal Strategies in Sports*, pages 77–85. Elsevier-North Holland, New York, 1977.

[2] Bukiet, B., E. R. Harold, and J. L. Palacios. A Markov chain approach to baseball. *Operations Research*, 25:14–23, 1997.

[3] Courneya, Kerry S. and Albert V. Carron. Batting First Versus Last: Implications for the Home Advantage. *Journal of Sport and Exercise Psychology*, 12:312–316, 1990.

[4] Courneya, Kerry S. and Albert V. Carron. The Home Advantage in Sport Competitions: A Literature Review. *Journal of Sport and Exercise Psychology*, 14:13–27, 1992.

[5] Fudenberg, D. and J. Tirole. *Game Theory*. MIT Press, Cambridge MA, 1991.

[6] Howard, R. A. *Dynamic programming and Markov processes*. MIT Press and Wiley, 1960.

[7] Lindsey, G. R. An investigation of strategies in baseball. *Operations Research*, 11:477–501, 1963.

[8] Nemec, D. *The Rules of Baseball: An Anecdotal Look at the Rules of Baseball and How They Came to Be*. The Lyons Press, 1994.

[9] Simon, Gary A. and Jeffrey S. Simonoff. “Last Licks”: Do They Really Help? *The American Statistician* 60:13–18, 2006.

[10] Trueman, R. E. Analysis of baseball as a Markov process. In S. P. Ladany and R. E. Machol, editors, *Optimal Strategies in Sports*, pages 63–76. Elsevier-North Holland, New York, 1977.

[11] Turocy, T. L. The Theory of Theft: An Inspection Game Model of the Stolen Base Play in Baseball. Typescript, Texas A&M University, January 2004.

[12] Wright, C. R. and T. House. *The Diamond Appraised*. New York: Simon and Schuster, 1989.