
Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

# Sticking with What (Barely) Worked: A Test of Outcome Bias

Lars Lefgren, Brennan Platt, Joseph Price

MANAGEMENT SCIENCE
Vol. 61, No. 5, May 2015, pp. 1121–1136
ISSN 0025-1909 (print) | ISSN 1526-5501 (online)

http://dx.doi.org/10.1287/mnsc.2014.1966




# Sticking with What (Barely) Worked: A Test of Outcome Bias

Lars Lefgren, Brennan Platt, Joseph Price
Department of Economics, Brigham Young University, Provo, Utah 84602
{lars_lefgren@byu.edu, brennan_platt@byu.edu, joseph_price@byu.edu}

Outcome bias occurs when an evaluator considers ex post outcomes when judging whether a choice was correct ex ante. We formalize this cognitive bias in a simple model of distorted Bayesian updating. We then examine strategy changes made by professional basketball coaches. We find that they are more likely to revise their strategy after a loss than a win—even for narrow losses, which are uninformative about team effectiveness. This increased strategy revision following a loss occurs even when a loss was expected and even when failure is due to factors beyond the team's control. These results are consistent with our model's predictions.

Data, as supplemental material, are available at http://dx.doi.org/10.1287/mnsc.2014.1966.

*Keywords*: outcome bias; hindsight bias; Bayesian updating; strategy revision
*History*: Received May 9, 2012; accepted March 14, 2014, by Yuval Rottenstreich, judgment and decision making. Published online in *Articles in Advance* July 21, 2014.

## 1. Introduction

In a broad variety of settings, economic actors must regularly evaluate whether their current strategy is still optimal in a constantly shifting environment. Firms adjust product and pricing decisions as technology and consumer preferences change. University faculty update their research strategy in response to changes in professional norms and realized successes. Parents revise rules and incentives as their child's needs and circumstances change.

These revisions of strategy crucially rely on the actor's ability to process information about which parts of his strategy are working well and which need adjustment; yet, in practice, these evaluations are not necessarily objective and dispassionate. When asking individuals to assess the appropriateness of an action, psychologists have documented an *outcome bias* in their evaluation: even when ex ante information is identical, the action is considered more justified if the ex post outcome was favorable. Baron and Hershey (1988) show that individuals were more likely to be critical of a medical decision when the outcome was poor, even though the objective risk of a poor outcome was the same. Thus, an outcome-biased actor is likely to revise strategies suboptimally—failing to make needed adjustments after fortuitous successes and changing excessively after unlucky failures.

Prior empirical research has documented an outcome bias in various experimental laboratory settings. We examine outcome bias in a high-stakes environment in which this bias could have a dramatic effect on an individual's career and earnings. We also provide a simple theoretical model of outcome bias, which produces behavior consistent with our empirical findings. In our model, actors primarily follow standard Bayesian updating but place inordinately more weight on realized success or failure than other available information about future success.

Empirical examinations of how economic actors adjust strategies are complicated by the fact that it can be difficult to describe a strategy in a parsimonious fashion. Additionally, this requires panel data on the strategy and its consequences over multiple periods. We overcome these challenges by examining how National Basketball Association (NBA) coaches adjust their strategies going into a game in response to past success or failure. This is a high-stakes setting in which coaches have a strong career incentive to implement a strategy each week that maximizes the probability of victory. Additionally, we have detailed data from 23,275 games over 20 seasons.

The strategic decision we focus on is which five players start the game on the court (referred to as *starters* for the game, or the *starting lineup*). This is an important and easily observed decision that coaches make prior to the start of every game. We find that a coach is 17% more likely to keep the same five starters following a win than following a loss. This simple finding can be readily explained by the fact that when the team wins, the coach receives a positive signal regarding the efficacy of his strategy. However, a closer examination of the empirical findings suggests the existence of outcome bias.

First, the persistence in the starting lineup remains higher after narrow victories versus narrow losses, even though either outcome has the same information content regarding future success. Second, a starting lineup that has occurred more often previously in the season shows greater persistence across games and exhibits less evidence of outcome bias. Third, coaches' decisions to change strategies are equally responsive to expected and unexpected success. If coaches are optimally incorporating information, they should respond only to unexpected performance. Fourth, coaches adjust their starting lineups in response to success or failure based on factors outside their control (such as whether or not the other team makes their free throws).[1]

To provide a clearer definition of outcome bias, we develop a simple model of an individual evaluating whether to switch between two strategies. After choosing one, he observes a noisy measure (on the real line) of the quality of his choice. His goal is to maximize the probability of success, which occurs if the measure lies above zero.[2] The individual observes the success or failure (the *outcome*) as well as the continuous measure that generates it (overall *performance*); he then uses Bayes' rule to estimate the expected future success of both options, switching if the other strategy is more likely to succeed.

We incorporate outcome bias by assuming that the individual inflates the ex ante likelihood of the outcome that actually occurred; that is, he acts as if success was more likely than it really was following a successful outcome. This creates a discontinuous jump in the probability of switching strategies when comparing performance just below or just above zero, whereas strategy revision will appear Bayesian for performance further from the threshold. This outcome bias can also induce an individual to switch strategies after failure as a result of events outside his control.

Our results suggest that NBA coaches exhibit outcome bias in that they attribute excess importance to the role of their strategy in determining a narrow win or loss. Consequently, they may switch strategies excessively after losses and not enough after wins. In this regard, outcome bias is subtle. The decision maker is not consistently overoptimistic; rather, he swings from self-assurance to second-guessing. He does not ignore information; rather, he double-counts it by considering performance and its coarser measure, outcome. Our findings suggest that outcome bias may make it difficult for economic agents to make optimal strategic choices in a variety of settings, most noticeably for decisions where the strategy barely worked (or barely failed).

## 2. Prior Literature

Evaluating decisions made under uncertainty is not an easy task; psychologists have documented a number of cognitive biases that can distort the evaluation.[3] For our current setting, the two most relevant are outcome bias and hindsight bias.

When judging the correctness of a decision, one typically evaluates it from the *ex ante* position of the decision maker, asking whether it was the best choice given the information available at the time. *Outcome bias* occurs when the evaluator considers the *ex post* outcome as well.[4] This bias was first labeled by Baron and Hershey (1988). In their study, students were given objective data on the risks of a medical procedure and were asked to rate the correctness of a decision. The students consistently felt the decision was more justified when the outcome was successful than when it failed, even though all other information was unchanged. Positive outcomes also produced a more favorable view of gambling decisions.

Similar outcome bias has been shown in a variety of laboratory settings, rating ethically questionable choices (Gino et al. 2008), decisions that benefit one while causing greater harm to another (Gino et al. 2010), decisions by a fictitious salesperson to pursue one client over another (Marshall and Mowen 1993), hypothetical military decisions (Lipshitz 1989), and decisions as whether to evacuate before a hurricane (Tinsley et al. 2012).

Ratner and Herbst (2005) take this a step further to consider how outcome bias affects future decisions. Students were given two investment options, one of which had a clearly higher expected return and was thus the initial choice of most students. After learning the realized returns, students showed outcome bias, rating their own decision more favorably when the outcome was positive. When revisiting the two options for a second round, 23% of those with bad outcomes switched to the lower average return option,

[1] Throughout our paper we refer to the coach as being the one exhibiting outcome bias. Coaches have a broad set of constituencies that they must appease—most notably, the owner, general manager, and fans. If these other groups are prone to outcome bias, then even a perfectly rational coach might appear to exhibit outcome bias (Denrell et al. 2013). This would not affect the analysis and results of our paper, just shift the source of the outcome bias from the coach to another group.

[2] This will be natural in our empirical setting. The quality of a team's performance may be best measured by the difference between its score and it opponent's, but the most salient measure of success is whether the team won or lost. Even so, this setup is relevant for a much broader variety of settings. For example, students are categorized as having failed a course based on their inability to cross a threshold level of performance. Managers may be considered unsuccessful if they do not hit predetermined targets for sales or profits. Doctors may be particularly concerned about whether a sick patient dies, despite other measures such as prolonging life or reducing pain.

[3] Earl (1990) and Rabin (1998) survey a much larger set of cognitive limitations, describing their potential importance in economic settings.

[4] If the outcome could reveal additional information held by the decision maker, the evaluator is not considered biased for considering it. For instance, if a home buyer discovers a massive mold problem shortly after closing, he may reasonably suspect the seller had more information than she disclosed.

whereas only 2% of those with good outcomes switched. Significantly, this demonstrates that outcome bias can occur in evaluating one's *own* decisions and can distort one's *future* decisions. This is also seen in hurricane evacuations, with residents being less likely to evacuate if minimal damage was sustained during the previous evacuation order under similar conditions (Tinsley et al. 2012).

A much larger literature considers *hindsight bias* (Fischhoff 1975), which occurs when people with knowledge of an outcome falsely believe they would have predicted that outcome. Of course, hindsight bias could easily lead to outcome bias: if the evaluator believes the outcome was inevitable, he will condemn the decision maker for not acting accordingly. Thus, some studies on hindsight bias overlap with the results on outcome bias. For example, LaBine and LaBine (1996) asked participants to act as jurors in a hypothetical malpractice suit for a therapist of a potentially violent patient. They assessed what the therapist should have known (hindsight bias was found) and should have done (outcome bias was found).

Hindsight bias has been studied extensively in the laboratory as well as in political polling; early studies were surveyed by Hawkins and Hastie (1990). The documented fact is that, after learning the outcome, people tend to shift their assessment of the expected outcome toward the true outcome. For instance, pollsters have asked voters on the day of an election what they expect the percentage outcomes to be and the next day asked voters to express what they thought the outcome would be. These recalled expectations are consistently closer to the actual result (Hawkins and Hastie 1990, p. 317).

Outcome and hindsight bias have received some limited attention from economists. Camerer et al. (1989) study whether more-informed participants in a market game can reproduce the judgments of participants with a given subset of information. Consistent with hindsight bias, the informed are swayed by their added knowledge of the outcome and tend to make worse decisions as a consequence. Outcome bias could also contribute to the *hot-hand effect* in sports, in which bettors overestimate the autocorrelation in performance. Offerman and Sonnemans (2004) introduce two distortions to standard Bayesian updating and determine that the hot-hand distortion can effectively explain their data, whereas the other distortion (recency) has no significant impact. Dobbs' (1991) purely theoretical study also has some relevance. He considers Bayesian updating in the face of ambiguity; that is, agents are unsure which probability distribution governs a random event, and they update their beliefs of the likely distribution after observing the outcome. After such updating, agents would appear to have hindsight bias: their best estimates of the ex ante probability of the event that occurred will always be higher than the same estimates before learning the outcome.

The current study offers several advantages over the preceding literature on outcome and hindsight bias. First, we study real-life decisions by experts in a high-stakes setting; our results confirm that outcome bias occurs outside the laboratory. Second, in our setting, coaches are evaluating their own decisions, rather than the decisions of others (as is common in most of the preceding literature).[5] It is easier to be critical or dismissive of the choices of others, and hence it is more significant to find outcome bias in self-evaluation. Third, we examine how these biases distort future decisions, which ultimately determines whether these biases actually matter. This issue is directly addressed only by Camerer et al. (1989), Ratner and Herbst (2005), and Tinsley et al. (2012). Fourth, we construct a model of decision making in which otherwise Bayesian agents exhibit outcome bias. In doing so, we translate a specific cognitive limitation into a set of theoretical predictions that are highly consistent with our empirical analysis.

Our project adds to a body of prior work in which sports competitions have provided a fertile setting for testing economic theories. For example, Pope and Schweitzer (2011) test for players' loss aversion in golf tournaments, and Card and Dahl (2011) test for loss aversion among National Football League (NFL) fans. Romer (2006) shows that football coaches fail to optimize when deciding whether to "go for it" on a fourth down. Chiappori et al. (2002) test whether soccer players behave as predicted by a simple "matching pennies" game when making penalty kicks. Gray and Gray (1997) test the efficiency of betting markets in the NFL, whereas Berger and Pope (2011) test prospect theory using the performance of NBA teams in the second half of a game based on their relative positions after the first half. Our work is similar in spirit to these earlier papers.[6]

Our paper also relates to a growing literature that examines whether economic agents are held accountable for factors outside of their control. For example, Bertrand and Mullainathan (2001) show that chief executive officers are rewarded on the basis of share price movements that are due to factors outside of the manager's control. Wolfers (2002) presents evidence that voters are less likely to vote for incumbent politicians during poor economic circumstances, even when the difficulties cannot be attributed to the politician. Collectively, these and other papers show that

[5] Jones et al. (1997), Ratner and Herbst (2005), and Gray et al. (2007) are notable exceptions. Also, one study by Tinsley et al. (2012) features residents evaluating their actual evacuation decisions, rather than hypothetical scenarios.

[6] In an earlier version of this paper, we examined the strategic choice of how frequently to pass the ball (rather than run it) in American football. We found similar evidence of outcome bias in each dimension. We prefer the current setting of the NBA because the binary nature of whether to change starters more closely resembles our theoretical model.

individuals have difficulty assessing other people's contributions to a successful or unsuccessful outcome. Our paper builds on this literature by demonstrating how people hold *themselves* accountable for success or failure unrelated to the quality of their own decisions.

## 3. A Model of Outcome Bias

We now present a simple model in which an agent evaluates the effectiveness of a particular strategy via Bayesian updating. We then expand the model to allow for an outcome bias.

### 3.1. Bayesian Updating

Consider an environment in which a single decision maker (a *coach*) must select either game plan $a$ or $b$.[7] In state $A$, the observed level of performance $P$ of plan $a$ is normally distributed with mean $h$ and variance $\sigma^2$, whereas plan $b$ produces the same variance but mean $l < h$. In state $B$, the means of the performance are reversed.[8]

Let $\rho \in [0, 1]$ denote the prior belief that $A$ is the current state, and let $\delta \in [0, 1/2)$ denote the probability that the state changes between periods. After an observed performance $P$, the coach uses Bayesian updating to determine the posterior probability that $A$ is the current state. For instance, if the coach used plan $a$ and observed performance $P$, the likelihood of this performance in states $A$ and $B$, respectively, would be

$$\Pr(P \mid A) = \frac{e^{-(1/(2\sigma^2))(P-h)^2}}{\sqrt{2\pi}\cdot\sigma} \quad \text{and} \quad \Pr(P \mid B) = \frac{e^{-(1/(2\sigma^2))(P-l)^2}}{\sqrt{2\pi}\cdot\sigma}.$$

Thus, the estimated posterior of the *Bayesian coach* would be

$$\begin{aligned}\hat{\rho} &= (1-\delta)\cdot\frac{\rho\cdot\Pr(P\mid A)}{\rho\cdot\Pr(P\mid A)+(1-\rho)\cdot\Pr(P\mid B)} + \delta\cdot\frac{(1-\rho)\cdot\Pr(P\mid B)}{\rho\cdot\Pr(P\mid A)+(1-\rho)\cdot\Pr(P\mid B)}\\ &= \delta - \frac{\rho\cdot(1-2\delta)}{\rho+(1-\rho)\cdot e^{(1/(2\sigma^2))(h-l)(h+l-2P)}}.\end{aligned}$$

[7] For our setting, this can be thought of as comparing the default starting lineup to the next best option. This can be easily generalized to many alternatives. For each alternative, there should be at least one state of the world in which it provides the best average payoff. Bayesian updating would occur across all states, with the team selecting the plan with the highest expected payoff.

[8] Here, the game plan does not specify each play, which needs to be randomized to avoid exploitation by the opponent. Rather, it should be seen as the broad strategy (such as which starting lineup will be most productive). The effectiveness of the game plan is mostly a question of how well it draws on the particular strengths of the team—a good fit will lead to higher average performance. In-game strategic decisions (such as choosing a specific play that the opponent does not anticipate) are reflected in the variance of performance.

We interpret the performance of a team's game plan as the difference between the final scores of that team and its opponent. Thus, if $P > 0$, the outcome of the game is that the team wins. If the coach's objective is to win as frequently as possible, he should employ plan $a$ as long as $\rho \geq 1/2$ and switch to plan $b$ otherwise.

By solving for the $P$ where $\hat{\rho} = 1/2$ for a given $\rho$, one obtains a performance threshold $\hat{P}(\rho) = (h+l)/2 + (\sigma^2/(h-l))\ln((1-\rho)/\rho)$. If performance falls below this threshold, the coach will switch plans before the next game. It is noteworthy that, generally, $\hat{P}(\rho) \neq 0$; that is, changing plans will not typically hinge on whether the team wins or loses. For example, if a coach has a strong prior that he is using the right strategy ($\rho$ is close to 1), the last term will be strongly negative. He could suffer a large loss and still remain convinced that he has the right strategy (though $\rho$ will fall after each poor performance, eventually leading to a change). Indeed, $\hat{P}(\rho) = 0$ only when $\rho = 1/(1 + e^{(h^2-l^2)/2\sigma^2})$.

### 3.2. Outcome Bias

To incorporate an outcome bias, we assume that the coach overweights the likelihood of the outcome that actually occurred by a factor $\gamma \geq 1$; that is, having used plan $a$, the coach interprets a winning (losing) outcome as additional evidence that $A$ ($B$) was the current state, even though performance $P$ fully accounts for that information. This distorts the ex ante likelihood used in Bayesian updating to make the observed win (loss) more likely than it really was. Formally, a *biased coach* using plan $a$ and observing outcome $P > 0$ obtains a posterior:

$$\begin{aligned}\check{\rho} &= (1-\delta)\cdot\frac{\rho\cdot\gamma\cdot\Pr(P\mid A)}{\rho\cdot\gamma\cdot\Pr(P\mid A)+(1-\rho)\cdot\Pr(P\mid B)} + \delta\cdot\frac{(1-\rho)\cdot\Pr(P\mid B)}{\rho\cdot\gamma\cdot\Pr(P\mid A)+(1-\rho)\cdot\Pr(P\mid B)}\\ &= \delta - \frac{\rho\cdot\gamma\cdot(1-2\delta)}{\rho\cdot\gamma+(1-\rho)\cdot e^{(1/(2\sigma^2))(h-l)(h+l-2P)}}.\end{aligned}$$

But using plan $a$ and observing outcome $P < 0$, the coach reaches a posterior:[9]

$$\begin{aligned}\check{\rho} &= (1-\delta)\cdot\frac{\rho\cdot\Pr(P\mid A)}{\rho\cdot\Pr(P\mid A)+(1-\rho)\cdot\gamma\cdot\Pr(P\mid B)} + \delta\cdot\frac{(1-\rho)\cdot\gamma\cdot\Pr(P\mid B)}{\rho\cdot\Pr(P\mid A)+(1-\rho)\cdot\gamma\cdot\Pr(P\mid B)}\\ &= \delta - \frac{\rho\cdot(1-2\delta)}{\rho+(1-\rho)\cdot\gamma\cdot e^{(1/(2\sigma^2))(h-l)(h+l-2P)}}.\end{aligned}$$

[9] The hot-hand distortion in Offerman and Sonnemans (2004) has some similarity to our model, in that agents overweight the ex ante probability that the coin is unfair when using Bayes' rule. The key difference is that our teams overweight the ex ante probability of whichever outcome (win/loss) actually occurred, rather than always overweighting the same event.

Performance $P=0$ almost never occurs, and hence it can be resolved either way without loss of generality. Note, however, that as $P$ increases, $\check{\rho}$ jumps upward discontinuously at $P=0$. This affects performance threshold $\check{P}(\rho)$, below which the coach switches plans:

$$\check{P}(\rho) = \begin{cases} \dfrac{h+l}{2}+\dfrac{\sigma^2}{h-l}\ln\left(\dfrac{1-\rho}{\rho\gamma}\right) & \text{if } \rho > \dfrac{\gamma}{\gamma+\mathrm{e}^{(h^2-l^2)/2\sigma^2}}, \\ \dfrac{h+l}{2}+\dfrac{\sigma^2}{h-l}\ln\left(\dfrac{(1-\rho)\gamma}{\rho}\right) & \text{if } \rho < \dfrac{1}{1+\gamma\mathrm{e}^{(h^2-l^2)/2\sigma^2}}, \\ 0 & \text{otherwise.} \end{cases}$$

With outcome bias, $\check{P}(\rho)$ equals 0 for an interval of priors, which we call $S$:

$$S \equiv \left[1/(1+\gamma\mathrm{e}^{(h^2-l^2)/2\sigma^2}),\ \gamma/(\gamma+\mathrm{e}^{(h^2-l^2)/2\sigma^2})\right].$$

Within this subset of priors, the coach will reach a posterior $\check{\rho} > 1/2$ for any $P>0$, and reach a posterior $\check{\rho} < 1/2$ for any $P<0$. Thus, if $\rho \in S$, then the coach switches strategies if and only if he loses the game. As $\gamma$ increases, the posterior increases for $P>0$ and decreases for $P<0$, making the size of the discontinuity (and the interval $S$) larger.

### 3.3. Consequences

We now present several predictions of this model that shed light on our empirical work.

*Prediction* 1. Under either Bayesian or biased updating, a coach is more likely to change strategies after worse performance. Yet, under biased updating, a coach is more likely to change strategies even when comparing a narrow loss to a narrow victory.

The first claim simply comes from $\partial\hat{\rho}/\partial P > 0$ and $\partial\check{\rho}/\partial P > 0$. A worse performance results in a lower posterior, and thus a coach is more likely to switch strategies. The second claim comes from the discontinuity at $P=0$. Any coaches who entered the game with $\rho \in S$ will switch strategies if and only if $P<0$. When $\gamma$ is moderately large, the interval $S$ will include the priors of many coaches. Thus, we will see that bias leads coaches (on average) to switch more often after losing than after winning, even when the margin of victory (or loss) is small. This would not show up for Bayesian coaches because a narrow loss only has a little more negative information (about future performance) than a narrow win and would thus only cause the tiny fraction of coaches with priors near $\rho = 1/(1+\mathrm{e}^{(h^2-l^2)/2\sigma^2})$ to change strategies.

It is worth noting that biased coaches are mistaken on both sides of $P=0$. They are too complacent after a narrow victory, overestimating $\check{\rho}$, and too worrisome after a narrow loss, underestimating $\check{\rho}$. These mistakes lessen as the magnitude of performance increases (such as decisive victories or defeats).

*Prediction* 2. Under either Bayesian or biased updating, a coach with a stronger prior is less likely to change strategies after a similar performance. Under biased updating, a coach with a stronger prior will appear to be less biased than one with a weaker prior.

Past performance is encapsulated in the prior belief $\rho$, which is higher after repeated success. This has two effects on updating. First, $\partial\hat{\rho}/\partial\rho > 0$ and $\partial\check{\rho}/\partial\rho > 0$, which is to say that with a stronger prior and the same realized performance, the posterior will also be stronger. Because of the accumulated positive evidence, the coach's confidence in his current strategy is less shaken by a given event. Thus, the posterior is less likely to fall below 1/2 and lead to a change of game plan.

A strong prior will also make outcome bias less visible, though. For instance, if $\rho > \gamma/(\gamma+\mathrm{e}^{(h^2-l^2)/2\sigma^2})$, then $\check{P}(\rho) < 0$; that is, barely losing is not enough to warrant a change in game plan. Indeed, the higher $\rho$ is, the worse the performance that will be tolerated without switching game plans. Yet even then, $\gamma$ still distorts the calculation of $\check{\rho}$, as the posterior still drops discontinuously as performance falls below $P=0$. Our point is merely that it will not fall below 1/2 for those with sufficiently high $\rho$. Empirically, coaches will appear to use standard Bayesian updating when they have more confidence in a particular strategy, even if they use the same biased updating process.

*Prediction* 3. Under Bayesian updating, only unexpected performance will affect the likelihood of changing game plans; higher expected performance has no effect. Under biased updating, unexpected performance also matters, but even expected losses can lead to game plan changes.

Higher expected performance could enter our model in a number of ways, but the simplest is to assume that both $h$ and $l$ increase by equal amounts; that is, average team performance increases whether the team is using the right game plan or the wrong one, with the difference between them remaining the same.

When a Bayesian coach has higher expectations, he will hold his team to a higher standard: for a given performance $P$, the posterior $\hat{\rho}$ is lower as $h$ and $l$ both increase. Since the team with higher expectations should be able to accomplish more, an unchanged performance gives the coach less confidence that he has the right game plan. In fact, it is only the *difference* between expected and realized performance that matters in Bayesian updating. If $h$, $l$, and $P$ each increase by $\varepsilon$, there is literally no change in $\hat{\rho}$. Thus, when expected and unexpected performance are included in the same regression, the former should have no impact on the likelihood of changing strategies, whereas the latter will be negatively correlated with strategy changes.[10]

[10] Alternative interpretations of increased expectations (e.g., $h$ increases by more than $l$) could potentially introduce some correlation

For a biased coach, higher expected performance operates identically on $\breve{\rho}$. However, this does not mean they only react to unexpected performance, as Bayesian coaches do. This is because $\partial\breve{\rho}/\partial\gamma > 0$ if and only if $P > 0$; bias increases the posterior for wins and lowers it for losses, regardless of expectations in $h$ and $l$. Indeed, in the extreme case (as $\gamma$ becomes very large), the only factor that coaches consider in retaining their game plan is whether they won or lost, regardless of whether the outcome was expected or not. Thus, even with moderate bias, coaches will be more likely to adjust their game plan when performance is lower, even (to some extent) when that performance was expected.

*Prediction* 4. Under Bayesian updating, a coach should not switch game plans as a result of events that influence the outcome but are unrelated to the plan. Under biased updating, an unrelated event may induce switching, particularly in close outcomes.

Many random events contribute to the final outcome of the game, some of which do not depend on the team's strategy. For instance, in basketball, the players of one team are not allowed to interfere while their opponent shoots free throws. A Bayesian coach would not allow random events, such as an opponent's free throw yield, to taint the evaluation of his strategy. Rather, the observed performance $P$ would be adjusted for events that are unaffected by his strategy before performing Bayesian updating. Thus, while the coach still responds to the adjusted performance measure as before, he will completely ignore irrelevant factors.

A coach with outcome bias can perform the same adjustment, but we assume that the coach is still biased by the actual outcome rather than the adjusted outcome. For example, if the team only lost because of its opponent's high yield on free throws, the team evaluates its overall strategy with an adjusted performance $P > 0$ but still places extra weight on the ex ante likelihood of state $B$. The resulting posterior $\breve{\rho}$ will be lower than what a Bayesian coach would conclude, and hence the coach could switch strategies even when his strategy-specific evidence is favorable. Of course, this bias will be most noticeable in close games, since factors that are irrelevant to the offensive performance are likely to still be relevant to the actual outcome.

## 4. Data

Before we move forward to our empirical analyses, it is crucial to identify our empirical measure of a coach's strategy. Although the set of strategies and choices that coaches make is extremely large, to make empirical headway, it is necessary to characterize a specific strategy that is both measurable and relevant. The measure we construct is whether or not the coach changes the set of starters between adjacent games. Each game, the coach must decide which five players will be on the court at the start of the game.

There are a number of reasons that the choice of which players to start the game is an important strategic coaching decision. First, a coach will generally want to start his best players at each position to spread their rest time over a longer period of time and maximize effective playing time from those players. Second, since other teams tend to start the game with their best players, a coach will also want to start his best group of players to diminish the effectiveness of the minutes played by the other team's top players. Third, a coach will want his best group of players on the court at the beginning of the game because it is easier to start the game in the lead and maintain the advantage than try to come back from behind later in the game.[11]

We use data on NBA team performance from the 1991–2010 seasons, which were obtained from http://basketballreference.com. We include only games from the regular season, which in nonstrike years includes 82 games. Our complete sample of data includes 46,550 team-game observations. To measure expected performance, we use the gambling spread for each game, which is publicly available prior to the game and is an excellent predictor of the realized score differential. We define unexpected performance as the difference between the actual score differential and the gambling spread.

Table 1 shows summary statistics for our sample. We see that between games, approximately 31% of teams change at least one player in their starting lineup. Although by construction teams win half of all games on average, the standard deviation of the winning rate over the season is 0.157. The standard deviation of the score differential between the reference team and its opponent is 13.41 points. The average score differential between the winner and loser is 10.87 points, with a standard deviation of 7.84. Both expected and unexpected performance, by construction, have a mean of 0. Expected performance is less variable than unexpected

between higher expectations and strategy changes. As the gap widens between the right plan and the wrong plan, the same amount of subpar performance (i.e., unexpected performance) is interpreted more harshly. Even then, this indirect effect of expected performance should be of second-order impact compared with the direct effect of unexpected performance.

[11] To provide some additional information about the starter decision, we use data compiled by Arcidiacono et al. (2013) for all regular season games during the 2006–2009 NBA season that lists the 10 players on the court during each possession of each game (five from each team). Among all the different combinations of players on the court during a game, the starting five play, on average, 14.3 minutes together (more than a quarter of the full game). The next most common combination of players spends only 7.9 minutes on the court together. On average, the starters spend 6.8 minutes on the court together at the start of a game before the first substitution occurs.

**Table 1** **Descriptive Statistics**

| Variable | Mean | SD |
|---|---|---|
| *Change starters* | 0.309 | 0.462 |
| *Win* | 0.500 | 0.500 |
| *Season win rate* | 0.500 | 0.157 |
| *Score differential* | 0 | 13.405 |
| *Absolute score differential* | 10.87 | 7.837 |
| *Expected performance* | 0 | 7.228 |
| *Unexpected performance* | 0 | 11.460 |

*Note.* The sample includes all regular season games during the NBA seasons and includes team-game observations.

performance, with a standard deviation of 7.23 points compared with 11.46 points.

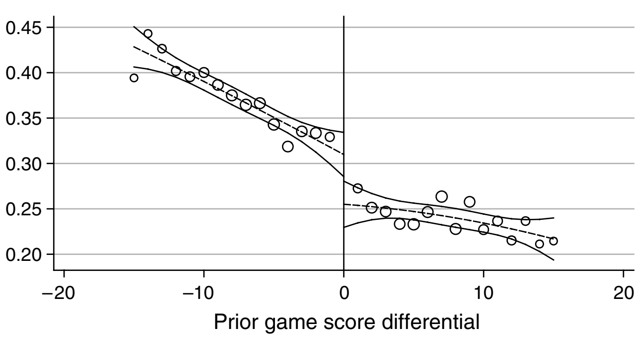
## 5. Empirical Evidence

### 5.1. Sticking with What Barely Worked

Our theoretical framework suggests that when coaches suffer from biased updating, their strategic decisions will be overly sensitive to the outcome of their last game. One manifestation of this is that even though team strategy affects a continuous performance measure (the score differential), coaches will ascribe additional importance to whether this performance resulted in a positive outcome (victory). Hence, coaches will be more likely to adjust their strategies after a narrow loss than after a narrow victory.

In Figure 1, we show graphical evidence consistent with this implication. Each circle represents the probability a coach changed starters (vertical axis) after a game that ended with a particular score differential (horizontal axis). The size of each circle is proportional to the number of games at each score differential. The dashed lines represent a second-order polynomial fit of the data, where the relationship is allowed to vary above and below the win threshold. The solid lines represent the 95% confidence intervals of the relationship. Although the probability of changing starters declines with the score differential, there is a discontinuous drop of approximately five percentage points at the win threshold. Unless winning the game by a small margin has strong information content regarding the performance of the team (which we examine later), this graphical evidence is suggestive that coaches suffer from outcome bias as they update their beliefs regarding the correct strategy.

**Figure 1** **Probability Starters Are Changed as a Function of Prior Game Score Differential**




*Notes.* The circles represent the mean outcome for each score differential value. The size of each circle is proportional to the number of games represented. The dashed line represents the predicted outcome, which was estimated with a second-order polynomial fully interacted with an indicator variable for whether the score differential was greater than or less than zero. The solid lines represent the 95% confidence intervals.

To test our theory more formally, we estimate the following regression discontinuity model:

$$Change\_starters_{i,g+1} = \beta_0 + \beta_1 Win_{i,g} + \beta_2 Score\_diff_{i,g} + \beta_3 Score\_diff_{i,g} \times Win_{i,g} + \varepsilon_{i,g},$$

where $i$ indexes the team, $g$ indexes the game during the season, $Change\_starters_{i,g+1}$ is a binary variable indicating whether the team changed starters for the next game, $Score\_diff_{i,g}$ is the score differential between the reference team and their opponent, $Win_{i,g}$ is an indicator variable of whether the team won or lost, and $\varepsilon_{i,g}$ is the error term. In this context, the score differential operates as the forcing variable that determines whether a team won the game. The interaction term between score differential and whether the team won allows the relationship between score differential and the probability of changing starters to vary above and below the win threshold. Unless indicated otherwise, in all specifications we cluster correct the standard errors at the team-season level. Consequently, this specification can be interpreted as a local linear regression with a rectangular kernel.

In Table 2, we show the empirical results from this estimation strategy. Column (1) shows the results with no additional covariates using only a five-point bandwidth around the win threshold, meaning that we include in this analysis only games that were decided by five or fewer points. Consistent with our graphical evidence, barely winning a game is associated with a 5.3-percentage-point drop in the probability that a team changes its starters (or 17% of the baseline probability). In column (2) we add team-season fixed effects to take into account any differences in team strength across the win threshold. The estimated impact of winning a game is slightly larger, at 5.8 percentage points. In columns (3) and (4) we increase the bandwidth to 10 and 15 points, respectively, and find virtually identical results.

In column (5) we control for whether the game was played at home and the winning percentage during the prior five games. The estimated coefficient on winning the game is a bit larger in absolute value but broadly consistent with the other results. The coefficient on the winning percentage in prior games is negative, suggesting that not only the current game but also

**Table 2** Regression Discontinuity Estimates of the Impact of Winning on the Probability of Changing Starters

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| *Win* | −0.053** | −0.058** | −0.052** | −0.052** | −0.061** |
| | (0.020) | (0.020) | (0.012) | (0.010) | (0.020) |
| *Score differential* | −0.001 | −0.001 | −0.008** | −0.007** | 0.001 |
| | (0.004) | (0.004) | (0.002) | (0.001) | (0.004) |
| *Score differential × Win* | −0.001 | −0.007 | 0.006** | 0.005** | −0.010* |
| | (0.006) | (0.006) | (0.002) | (0.001) | (0.060) |
| *Winning percentage in five games prior* | — | — | — | — | −0.155** |
| | | | | | (0.024) |
| *Home game* | — | — | — | — | 0.033** |
| | | | | | (0.008) |
| Team-season fixed effects | No | Yes | Yes | Yes | Yes |
| Bandwidth | 5 points | 5 points | 10 points | 15 points | 5 points |
| Observations | 12,972 | 12,972 | 26,565 | 35,309 | 12,160 |
| $R^2$ | 0.010 | 0.010 | 0.085 | 0.087 | 0.106 |

*Note.* Standard errors are in parentheses and are cluster corrected at the team-season level.

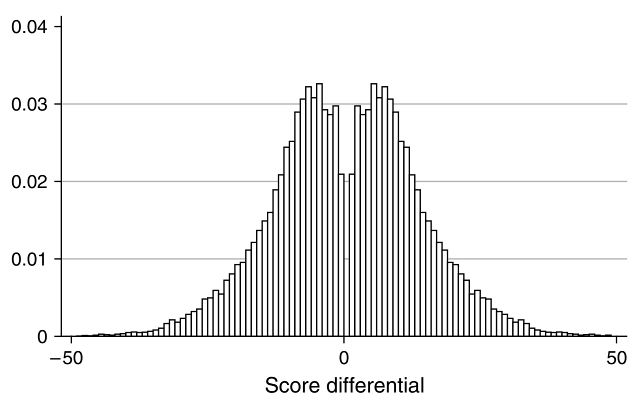
*Statistically significant at the 10% level; **statistically significant at the 5% level.

recent games affect the probability that a team changes starters.

The validity of the estimates from this regression discontinuity design depends on the comparability of teams on either side of the win cutoff. Traditionally, in regression discontinuity analyses, researchers test this assumption by looking for unusual clumping in the density of observations on either side of the cutoff and by examining whether the observable characteristics are balanced across the cutoff.

Figure 2 shows a histogram of games associated with each score differential. One notices a few peculiar characteristics of the histogram. The first is that there is no density at the cutoff itself. This is because games that are tied at the end of the regular playing period extend into overtime periods until one team has won.

**Figure 2** Histogram of Score Differentials




*Note.* The figure does not reflect 30 observations with a score differential of more than 50 points.

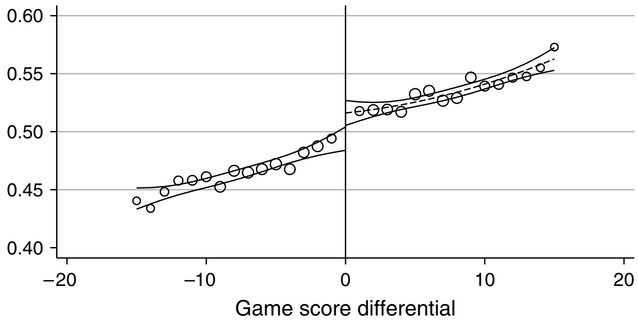
The density that would be at zero is spread across other score differentials slightly farther away from the threshold. The second observation is that the histogram is perfectly symmetric. This is driven by the fact that for each team that won by a particular point spread, there is a corresponding team that lost by the same spread. This symmetry mechanically ensures that there is no bunching in the density to one side of the cutoff, potentially alleviating concerns of the type raised by McCrary (2008). Third, there are fewer wins by one point than there are by two. This is likely driven by end-of-game incentives, with teams implementing strategies that make it difficult for their opponents to win the game with a single basket.

Although the histogram and institutions of the game make it clear that the absolute number of teams on either side of the cutoff will be exactly the same, it is unclear whether the teams on either side of the cutoff are comparable. In Figure 3, we look at the winning percentages of teams prior to the reference game.[12] We see that teams that just barely won a game are, in fact, stronger than teams that just barely lost, winning past games two to three percentage points more often. This suggests that teams just above and below the cutoff are not, in fact, perfectly comparable. However, if we examine a given team in a given season, this gap disappears; there is no information regarding the quality of play indicated by a close win or loss. To demonstrate this, we plot the win rate of the prior five games versus the score differential of the reference game, controlling for team fixed effects. Hence the outcome is the deviation in the winning percentage of the prior five games relative to the winning percentage of all other five game stretches across the season for the team in question. If narrow wins are informative of a good strategy, independent of team strength, we might expect the team to have performed well just prior to the narrow win. Instead, Figure 4 suggests that there is no significant difference between performance just above and just below the cutoff.[13] This evidence suggests that conditional upon overall team strength, there is little information in a narrow win as opposed to a narrow loss.

[12] The results are similar if we examine subsequent games, but we omit these results because the outcome of later games may be endogenous to the decision of changing starters.

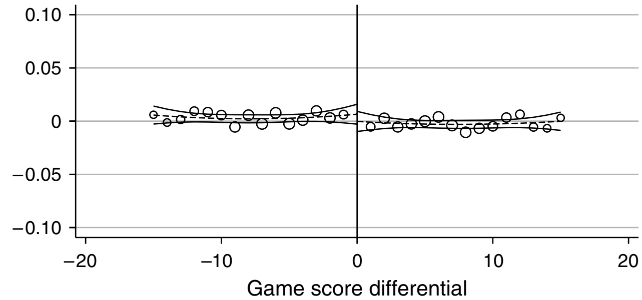
[13] The fixed effects specification has a negative downward bias due to an effect similar to a lagged dependent variable bias. This is because if the current game is a win, and hence above the team average win rate, all other games will be slightly more likely to have been losses. The bias is proportional to the inverse of the number of games in a season, which in most years is 82. Hence the bias is quite small. The bias does account, however, for the slight drop in prior win percentage above the win threshold. If, instead of controlling for team-season fixed effects, one controls for season winning percentage omitting the current game, the impact of winning a close game is still both small and statistically insignificant.

**Figure 3** **Winning Percentage in Prior Games as a Function of Prior Game Score Differential**




*Notes.* The circles represent the mean outcome for each score differential value. The size of each circle is proportional to the number of games represented. The dashed line represents the predicted outcome, which was estimated with a second-order polynomial fully interacted with an indicator variable for whether the score differential was greater than or less than zero. The solid lines represent the 95% confidence intervals.

**Figure 4** **Winning Percentage Residual in Five Prior Games as a Function of Prior Game Score Differential**




*Notes.* The dependent variable is the residual of a regression of the winning percentage in five prior games on team-season fixed effects. The circles represent the mean outcome for each score differential value. The size of each circle is proportional to the number of games represented. The dashed line represents the predicted outcome, which was estimated with a second-order polynomial fully interacted with an indicator variable for whether the score differential was greater than or less than zero. The solid lines represent the 95% confidence intervals.

In Table 3, we employ the same type of regression discontinuity (RD)-specification to more formally examine the relationship between a narrow win and prior success. Except for the narrowest bandwidth of five points, a narrow win is associated with a higher winning percentage in prior games of the season, consistent with Figure 3. In columns (4)–(6), however, we see that a narrow win is uncorrelated with the win percentage in the prior five games once we control for team-season fixed effects, as in Figure 4. This provides more formal evidence that, conditional on team strength (as captured by the team-season fixed effects), a narrow win is uninformative regarding the efficacy of the coach's strategy.

In summary, the evidence suggests that teams do vary across the win threshold on the basis of strength, casting some doubt on the interpretation of our RD estimates of the impact of a narrow win on the probability of a coach changing his strategy. However, conditional upon team fixed effects, it does not appear that a narrow win contains information about how the team is playing during the period of the season in question. Hence, an alternative identification assumption may hold that for a given team during a season, a narrow win is not informative about how the team is playing relative to a narrow loss. Hence, coaches should not make strategic decisions based on narrow victories or losses. Note that the estimates shown in columns (2)–(5) of Table 2 include team-season fixed effects, consistent with this assumption, yet the estimates indicate that coaches are heavily influenced by whether they won or lost.

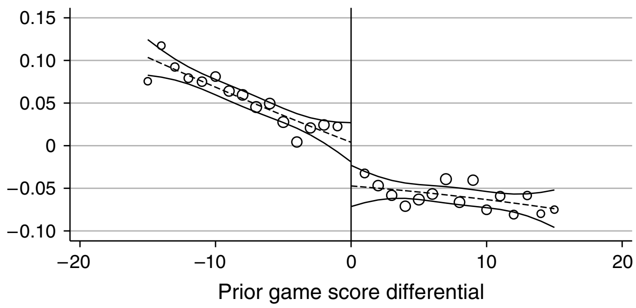
In Figure 5 we show additional graphical evidence consistent with this alternative identification assumption. In particular, we plot residuals of a regression of an indicator of whether the team changed starters on team-season fixed effects. Hence, the figure graphs the probability of changing starters minus each team's season-long propensity to change starters against the score differential of the current game. Although the

**Table 3** **Regression Discontinuity Estimates of the Impact of Winning on the Winning Percentage in Prior Games**

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | Win % all prior games | | | Win % prior five games | | |
| *Win* | 0.012 | 0.020** | 0.020** | −0.012 | −0.004 | −0.006 |
| | (0.008) | (0.005) | (0.005) | (0.009) | (0.005) | (0.004) |
| *Score differential* | 0.006** | 0.004** | 0.003** | 0.002 | 0.000 | −0.000 |
| | (0.002) | (0.001) | (0.000) | (0.002) | (0.001) | (0.000) |
| *Score differential* × *Win* | −0.003 | −0.001 | −0.000 | −0.002 | −0.001 | 0.000 |
| | (0.002) | (0.001) | (0.001) | (0.003) | (0.001) | (0.001) |
| Team-season fixed effects | No | No | No | Yes | Yes | Yes |
| Bandwidth | 5 points | 10 points | 15 points | 5 points | 10 points | 15 points |
| Observations | 12,977 | 26,551 | 35,262 | 12,324 | 25,198 | 33,494 |
| $R^2$ | 0.014 | 0.026 | 0.038 | 0.411 | 0.405 | 0.407 |

*Note.* Standard errors are in parentheses and are cluster corrected at the team-season level.
**Statistically significant at the 5% level.

**Figure 5** **Probability Starters Are Changed Residual as a Function of Prior Game Score Differential**




*Notes.* The dependent variable is the residual of a regression of the winning percentage in five prior games on team-season fixed effects. The circles represent the mean outcome for each score differential value. The size of each circle is proportional to the number of games represented. The dashed line represents the predicted outcome, which was estimated with a second-order polynomial fully interacted with an indicator variable for whether the score differential was greater than or less than zero. The solid lines represent the 95% confidence intervals.

$y$ axis has shifted, the picture is virtually equivalent to Figure 1, suggesting the drop at the win threshold is not driven by discontinuous differences in team strength.

In Table 4, we examine a number of alternative specifications to examine the robustness of our finding that coaches are excessively sensitive to narrow wins in their strategy choice. In the first two rows we replace our linear specification with higher-order polynomials for the score differential. We use a larger bandwidth as we increase the order of the polynomial of the score differential to maintain the statistical power of our estimates. The results are similar to our baseline estimates of the impact. In the third row we examine only games in which all starters also played in the subsequent game. This specification is included to ensure that our results are not driven by injuries to the starters.[14] In this specification, the coefficient on winning is about 30% smaller in absolute value, but still statistically significant. The slight attenuation of the coefficient is likely because we are excluding instances in which the coach deliberately changed his strategy by completely excluding a starter from playing in the subsequent game.

In the final specification of Table 4, we construct a falsification exercise in which we examine whether a narrow victory affects the probability that the starters were changed in the previous game. If our results were driven by some omitted variable bias, we would expect the bias to operate both going forward as well as backward in time. Instead, we see that a narrow victory has no significant relationship with prior changes in strategy.[15]

**Table 4** **Regression Discontinuity Estimates of the Impact of Winning on the Probability of Changing Starters—Robustness Checks**

| | Bandwidth | Win coefficient |
|---|---|---|
| Alternative specification | | |
| Second-order polynomial in forcing variable | 10 points | −0.062** |
| | | (0.023) |
| Third-order polynomial in forcing variable | 15 points | −0.064** |
| | | (0.026) |
| Only games in which all starters play in next game | 5 points | −0.044** |
| | | (0.018) |
| Falsification test | | |
| Impact of close win on changing starters in previous game | 5 points | 0.018 |
| | | (0.020) |

*Notes.* Standard errors are in parentheses and are cluster corrected at the team-season level. All specifications include team-season fixed effects.
**Statistically significant at the 5% level.

### 5.2. Rationality of the Usual

The second prediction of our model is that a coach who is more certain of the appropriateness of a given strategy will engage in less switching based on the outcome of a particular game. To test this hypothesis, we examine whether the sensitivity of the starting lineup depends on how often the same players have started games in the past. We expect that coaches are more confident in using a particular set of players if the coach has used the same set often in the past.[16] In these situations, our model predicts that coaches will appear to have less outcome bias in their strategic decisions. To test this implication, we estimate the following model:

$$\begin{aligned} \mathit{Change\_starters}_{i,g+1} &= \beta_0 + \beta_1\, \mathit{Win}_{i,g} + \beta_2\, \mathit{Win}_{i,g} \times \mathit{Frac\_starts}_{i,g} \\ &\quad + \beta_3\, \mathit{Frac\_starts}_{i,g} + \beta_4\, \mathit{Score\_diff}_{i,g} \end{aligned}$$

[14] In 49.5% of cases in which a group of starters is changed, at least one of the players does not play at all in the subsequent game. This provides an upper bound to the fraction of times in which the starting lineup is changed because of injury.

[15] We also examined whether there was substantial heterogeneity in the degree of outcome bias across coaches or situations. Focusing on games late in the season, we did not find that teams exhibited significantly more or less outcome bias depending on whether they were likely in the playoffs, likely out of the playoffs, or on the margin of making it into the playoffs. To examine heterogeneity across coaches, we estimated the degree of outcome bias for each team-season combination. We then regressed our measure of team-season outcome bias on coach fixed effects. We could not reject the null hypothesis that coaches did not differ on the basis of outcome bias. We also found no relationship between experience and degree of outcome bias.

[16] One might wish to use the winning percentage of the prior lineup as an objective measure of the coach's prior of the starting lineup's effectiveness. We have done so, and although the point estimates are consistent with coaches exhibiting less outcome bias when the winning percentage of the starting lineup is higher, the estimates are not statistically significant. We choose to emphasize the results focusing on the fraction of higher starts because this is a revealed preference measure of the coach's belief of the lineup's effectiveness, even if the belief itself may be subject to cognitive biases.

$$+\beta_5\, Score\_diff_{i,g} \times Frac\_starts_{i,g}$$
$$+\beta_6\, Score\_diff_{i,g} \times Win_{i,g}$$
$$+\beta_7 Score\_diff_{i,g} \times Win_{i,g}$$
$$\times Frac\_starts_{i,g} + \varepsilon_{i,g}.$$

Note that $Frac\_starts_{i,g}$ is the fraction of previous games this season prior to game $g$ for which those five players were the starters for team $i$. This specification corresponds to an RD model in which both the effect of winning and the impact of score differential are allowed to vary linearly with the fraction of starts.

Table 5 shows the results from this estimation. We include only those games played after at least 20 games into the season so that coaches have time to establish a pattern of starters. The coefficient on the win variable indicates that for a team that has never started a particular group of players in the past, a narrow win reduces the probability that the starters are changed by about 10 percentage points. The coefficient on the interaction term shows that as the fraction of games those five players have started together previously grows, the sensitivity of the coach's decision to change the starters in response to a narrow loss rapidly declines. The coefficients on the direct effect of winning a game as well as the interaction effect of winning and the fraction of starts are all statistically significant at least at the 10% level, regardless of the bandwidth choice or the inclusion of team fixed effects and controls for other covariates. Using column (2) as our baseline suggests that for a set of players in the 25th percentile of prior starts (8% of games), winning a close game reduces the probability the starters are changed by 10.5 percentage points. For a set of players in the 75th percentile of prior starts (57% of games), winning a close game reduces the probability of changing starters by only 2.7 percentage points. Collectively, the results presented in this table provide strong evidence consistent with the implication of our theory that biased behavior will be less evident when coaches have a strong prior regarding the right strategy.

**Table 5** **Regression Discontinuity Estimates of the Impact of Winning on the Probability of Changing Starters by Fraction of Prior Starts**

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| *Win* | −0.105** | −0.118** | −0.109** | −0.074** | −0.115** |
| | (0.037) | (0.037) | (0.023) | (0.019) | (0.037) |
| *Win* × *Fraction starts* | 0.140* | 0.160** | 0.127** | 0.069* | 0.147* |
| | (0.079) | (0.080) | (0.050) | (0.040) | (0.081) |
| *Fraction starts* | −0.367** | −0.342** | −0.268** | −0.247** | −0.333** |
| | (0.055) | (0.058) | (0.040) | (0.034) | (0.058) |
| Team-season fixed effects | No | Yes | Yes | Yes | Yes |
| Other controls | No | No | No | No | Yes |
| Bandwidth | 5 points | 5 points | 10 points | 15 points | 5 points |
| Observations | 8,390 | 8,390 | 17,050 | 22,691 | 8,390 |
| $R^2$ | 0.037 | 0.130 | 0.105 | 0.101 | 0.133 |

*Notes.* *Fraction starts* represents the fraction of prior games in the season in which the current starting players also started. We only include games occurring after the first 20 games of the season. All specifications include linear terms for score differential and score differential interacted with whether the team won. We also interact these measures with *Fraction starts*. In specification (5) we also control for whether the game was at home and the winning percentage in the prior five games. Standard errors are in parentheses and are cluster corrected at the team-season level.

*Statistically significant at the 10% level; **statistically significant at the 5% level.

### 5.3. Reacting to the Expected

The third prediction of our model is that Bayesian coaches should update their strategies based only on performance that deviates from what was expected. For instance, when teams unexpectedly lose to a weak team, they should be more likely to update their strategy than when they lose to a strong team. However, both expected and unexpected performance will influence whether the team ultimately wins or loses. Consequently, coaches suffering from outcome bias will update their strategies both on the basis of expected and unexpected performance. We quantify expected performance by looking at the gambling spread. Unexpected performance is the difference between the actual point differential and the gambling spread.

More concretely, we estimate the following regression equation:

$$Change\_starters_{i,g+1} = \beta_0 + \beta_1\, Expected\_perf_{i,g} + \beta_2\, Unexpected\_perf_{i,g} + \varepsilon_{i,g},$$

where $Expected\_perf_{i,g}$ is the expected performance of team $i$ in game $g$, and $Unexpected\_perf_{i,g}$ represents the unexpected performance. If coaches use standard Bayesian updating, then only $\beta_2$ should be negative and significant. With outcome bias, however, we might expect both $\beta_1$ and $\beta_2$ to be negative and significant. In Table 6, we find that expected and unexpected performance affect the persistence of the coach's strategy in virtually identical ways. In fact, in column (1), where we do not control for team-season fixed effects, one cannot reject that teams are equally sensitive to both expected and unexpected performance. Once we include team-season fixed effects, teams respond slightly less, but still significantly, to the expected score differential. To put the magnitudes in perspective, if a team plays a weak opponent and, as expected, scores a net of 10 more points, the starters are five percentage points less likely to be changed. Yet if a team plays a difficult opponent and unexpectedly scores a net of 10 more points (relative to the point spread), the starters are six percentage points less likely to be changed. This table is strongly consistent with coaches being sensitive to the outcome of the game as opposed to the information content in the game.

**Table 6** Impact of Expected and Unexpected Performance on the Probability of Changing Starters

| | (1) | (2) |
|---|---|---|
| *Expected score differential* | −0.006** | −0.005** |
| | (0.000) | (0.000) |
| *Unexpected score differential* | −0.006** | −0.006** |
| | (0.000) | (0.000) |
| *F*-statistic of difference | 0.14 | 6.00 |
| *p*-value | 0.71 | 0.01 |
| Team-season fixed effects | No | Yes |
| Observations | 45,217 | 46,550 |
| $R^2$ | 0.028 | 0.011 |

*Note.* Standard errors are in parentheses and are cluster corrected at the team-season level.
**Statistically significant at the 5% level.

### 5.4. The Relevance of Irrelevant Information

Factors that are irrelevant to a team's strategy can affect whether a team ultimately wins; our fourth prediction says a biased coach may allow such events to taint the evaluation of his strategy. For example, a team has little control of how well their opponent shoots free throws. A free throw is an opportunity for a player to shoot at the basket from 15 feet away without any interference from opposing players. A free throw attempt is awarded to an opponent as a result of the team violating a rule of the game. On average, a team is awarded 25.5 free throws a game and converts on 75% of attempts. Because the opponent's free throw percentage is almost entirely outside of a team's control, a team should not make strategic decisions for the next game based on whether the opponent completed many or few of their free throw attempts. Table 7 shows regression results examining the impact of the opponent free throw conversion rate on a variety of outcomes. We control for no additional covariates except season fixed effects or team-season fixed effects. Free throw conversion rates have trended upward across seasons, which is why we control for season in all specifications.

In the first column, we see that opponent free throw percentage in the current game has absolutely no predictive power regarding the next opponent's free throw percentage in the subsequent game. This provides evidence that a team has virtually no influence on its opponents' free throw percentage. In columns (2) and (3) we show that opponent free throw percentage is an important predictor of the outcome of the current game.[17] In columns (4) and (5) we see that free throw percentage also has a significant effect on the probability that the starters have changed. If an opponent makes 80% of their free throws instead of 70%, the starters are about half a percentage point more likely to be changed in the next game. This table provides strong evidence in support of our final implication that biased coaches react to information that is irrelevant for understanding the expected efficacy of their strategy.

It might be the case that savvy teams are more likely to foul opponents who are poor free throw shooters. If this was an indicator of a strong team, however, we would expect opponent free throw shooting to be persistent over time. The results in column (1) of Table 7 indicate that this is not the case. In contrast, the number of free throws that an opponent shoots is correlated across games, indicating that some teams are better able to reduce the fouls they commit by allocating less time to foul-prone players or putting in better defenders. If we include the number of fouls that the opponent shoots as an additional control in Table 7, all of the results are nearly identical, providing additional evidence that a coach has almost no control over an opponent's free throw percentage.

[17] One approach would be to use opponent free throw conversion rate as an instrument for victory. We do not take this approach because it affects both the probability of victory and the margin of victory.

**Table 7** Impact of Opponent Free Throw Percentage on the Probability of Victory and Changing Starters for the Next Game

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | Next game opponent free throw percentage | Win current game | | Change starters | |
| Opponent free throw percentage coefficient | −0.000 | −0.537** | −0.520** | 0.053** | 0.066** |
| | (0.005) | (0.024) | (0.023) | (0.023) | (0.022) |
| Season fixed effects | Yes | Yes | No | Yes | No |
| Team × Season fixed effects | No | No | Yes | No | Yes |
| Observations | 45,971 | 46,550 | 46,550 | 45,971 | 45,971 |
| $R^2$ | 0.011 | 0.011 | 0.098 | 0.003 | 0.000 |

*Note.* Standard errors are in parentheses and are cluster corrected at the team-season level.
**Statistically significant at the 5% level.

## 6. Conclusion

Decision makers have difficulty evaluating the efficacy of a strategy when random events also influence the final outcome. One can easily misinterpret a favorable outcome as justification for a given strategy, overriding more subtle evidence to the contrary. In this paper, we provide a theoretical definition of outcome bias, which distorts Bayesian updating by overweighting the ex ante likelihood of the outcome that ex post occurred. This theory provides four clear predictions, all of which are borne out in our empirical application to the starter strategy of coaches in the NBA.

In particular, coaches tend to change their strategy more frequently after losing a game than after winning the game. This occurs even when comparing narrow losses or victories, where winning contains no information regarding the quality of team play. We also find that coaches are much less sensitive to a loss in situations where they are more certain about the appropriateness of a given strategy, they react equally to expected and unexpected performance, and their strategy is as likely to be revised whether or not responsibility for a loss is based on factors outside their control (such as whether the other team makes their free throws). All of these results are consistent with our theory of outcome bias.

It is not surprising that coaches would focus on the binary outcome of winning or losing a game; after all, this is what matters most to the team owners and fans, and it will largely determine whether the coach retains his job. But to maximize the chance of future wins, a coach ought to base his strategy revisions on information that most accurately predicts future success. Outcome bias reduces the accuracy of his judgments, leading to complacency after narrow wins and excessive switching after narrow losses.

To quantify the impact of bias on team behavior and game outcomes, we calibrate our biased model to match the data using the procedure reported in the appendix. One virtue of formal modeling is that it allows us to perform counterfactual experiments not seen in the data. In particular, we ask how a team would fare if its coach were free of outcome bias; that is, we set the bias parameter $\gamma = 1$, while holding all others constant, and simulate the outcome of 46,000 games. We find that starters would be changed after only 26.1% of the games, rather than 30.9% as in the data. Excessive switching could impose direct costs, taking an emotional toll on the players involved. The effect on game outcomes is more muted, reversing a win or loss in 4.8% of the games; on net, a Bayesian coach would win 0.7% more often.

Although the effect size that we calculate might be small, it is not far out of line with the findings in Price and Wolfers (2010) on racial bias among referees. They compute that if the racial composition of referees was adjusted to match the racial composition of the players, it would change the outcome of 1.6% of games. Moreover, our analysis examines just one of the many decisions that coaches make; the others may be similarly biased. As such, our results are likely to provide a conservative estimate of the overall impact of outcome bias on the full set of coaching decisions.

We chose to document outcome bias in this sports setting because of the ease of quantifying strategies, the availability of uniform data, and the high incentives for effective evaluation. However, we anticipate that decision makers in many other settings are equally susceptible to outcome bias. For instance, sales personnel are often judged relative to sales goals. A manager could easily place excessive weight on whether the salesperson cleared the mark, even when those just above the threshold may differ only in good fortune from those just below. If so, the manager would make inefficient decisions in revamping sales incentives or retaining employees. Good programs or workers (with a modest amount of bad luck) would be scrapped, whereas those that barely cleared the threshold would be given too much credit for their success.

## Supplemental Material

Supplemental material to this paper is available at http://dx.doi.org/10.1287/mnsc.2014.1966.

## Acknowledgments

The authors are grateful for excellent research assistance from Jason Cook, Craig Paulson, and Michael Worley. They also thank Dan Stone and seminar participants at Brigham Young University; the University of Nevada, Reno; Australia National University; and the 2013 Southern Economic Association Meetings for helpful comments. An earlier version of this paper was circulated under the title "Sticking with What (Barely) Worked" (NBER Working Paper 17477). Funding for this project was kindly provided by the College of Family Home and Social Science of Brigham Young University.

## Appendix. Calibration

To assess the impact of outcome bias, the ideal experiment would provide us with a biased coach compared to an otherwise identical unbiased coach. However, the data do not reveal some coaches as being more biased than others (see Footnote 15), and any variation that we do observe (such as frequent starters being less likely to be penalized for a loss) is still consistent with all coaches having the same bias but different priors (our second model prediction and empirical test). Thus, rather than using a natural experiment, we use the model to provide an estimate of this effect. Specifically, we calibrate the model parameters such that simulated data will replicate the statistics on score differentials and the probability of changing starters in our NBA data. Then, we change the bias parameter to $\gamma = 1$ and resimulate to determine how strategy changes and game outcomes would differ if outcome bias were absent.

We augment the model here with one feature: after each game, with probability $\tau$, the coach is perfectly informed of the need to change strategies for exogenous reasons (such as player injury, trades, or opponent matchups). This informed decision is distinct from the uncertainty studied in our model—affecting none of our conclusions—and is distinct from unobserved changes to the environment that occur with probability $\delta$. Even when a coach knows he has to remove one player, the remaining lineup is subject to the same complexities as before; thus he holds the same confidence (i.e., updated posterior) in his strategy choice after the exogenous change. These exogenous strategy changes are needed to replicate the fact that coaches sometimes make adjustments even when the team has been performing well.

The simulated data (for a given set of parameters) are generated from the perspective of one coach who repeatedly observes random performance $P$ and decides whether to change strategies, with changes recorded as $S = 1$ for both exogenous and elective strategy changes, and $S=0$ otherwise. Formally, this process iterates as follows:

1. We begin iteration $t$ given an unbiased probability of having the right strategy, $\rho_{u,t}$, and biased prior beliefs, $\rho_{b,t}$.
2. We draw a random variable $c_t \sim \mathrm{U}(0,1)$. If $c_t < \rho_{u,t}$, then we draw a random variable $P_{t|h} \sim \mathrm{N}(h, \sigma)$; otherwise, we draw $P_{t|l} \sim \mathrm{N}(l, \sigma)$. A win $W_t = 1$ is recorded iff $P_t > 0$.
3. We compute posterior $\hat{\rho}_t$ (the unbiased updated probability) as detailed in §2 using $\rho_{u,t}$ as the prior. We also compute posterior $\breve{\rho}_t$ (the biased updated belief) using $\rho_{b,t}$ as the prior.
4. If $\breve{\rho}_t < 1/2$, the biased coach will switch strategies, so $S_t = 1$. This interchanges the two strategies and inverts the probabilities; that is, the coach believes the newly selected strategy is correct with probability $\rho_{b,t+1} = 1 - \breve{\rho}_t$, whereas the correct probability is $\rho_{u,t+1} = 1 - \hat{\rho}_t$.
5. If $\breve{\rho}_t \geq 1/2$, we draw a random variable $z_t \sim \mathrm{U}(0,1)$. If $z_t < \tau$, the coach adjusts strategies for exogenous reasons, so $S_t = 1$. Otherwise, the strategy is unchanged, so $S_t = 0$. In either case, the coach moves forward with belief $\rho_{b,t+1} = \breve{\rho}_t$, although the true probability is $\rho_{u,t+1} = \hat{\rho}_t$.

Note that the unbiased true probability $\rho_{u,t}$ governs whether performance $P_t$ is drawn from the normal distribution with low or high mean, but the biased prior $\rho_{b,t}$ governs how that performance is interpreted and whether the strategy is changed. Also, as in the model, we are only considering the decision of one coach, rather than pairs of teams, in each contest.

Moving to the calibration process, we face a challenge in that our data give no indication of a coach's prior belief ($\rho$) of whether he is using the best set of starters. Of course, this prior evolves over time according to Bayesian updating and hence can take many different values over the course of even one season of 80 games. To sidestep this issue, we make an initial guess applied to both $\rho_{u,1}$ and $\rho_{b,1}$, and then we simulate a sequence of 46,000 games under a given set of parameters. By compiling $\rho_{b,t}$ across all iterations, we obtain a steady-state distribution of priors $\rho$ that is not sensitive to the initial guess for $\rho$, which we then take to represent the distribution of biased prior beliefs across teams. Figure A.1 provides a histogram of this distribution from our final calibration.

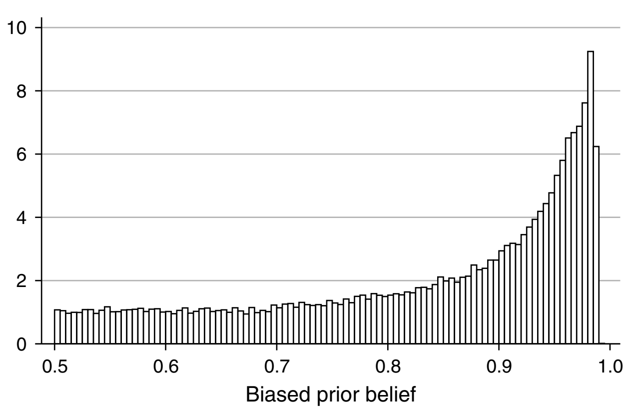
In Table A.1, we identify five moments of the NBA data (across all teams and seasons) chosen as calibration targets; all but the first are drawn directly from Table 1. Table A.1 also lists the remaining parameters, whose values are revised until the moments of the simulated data match the targets from the NBA data. Adjustment of one parameter can potentially affect all the targeted moments; in practice, however, the parameter has the strongest effect on the target by which it is listed in the table. For instance, $\tau$ directly controls the fraction of strategy changes following a win, whereas the standard deviation of realized performance is closely related to $\sigma$. This allows us to conduct the calibration sequentially (in the order listed), adjusting one parameter to meet its target and then moving to the next; this repeats until the process converges. The resulting parameter values are listed in Table A.1.

As a visual illustration of how well the simulated data mimics our NBA data, in Figure A.2 we recreate Figure 1 using data simulated under the calibrated parameters. Note that we are matching the aggregate moments, rather than attempting to do so on a team-by-team basis. From a practical standpoint, the calibration process relies on a very large sample size as a result of the noisiness of performance. For instance, with only a thousand observations, the moments of the simulated data can fluctuate 5%–10% between each

**Table A.1 Calibration Results**

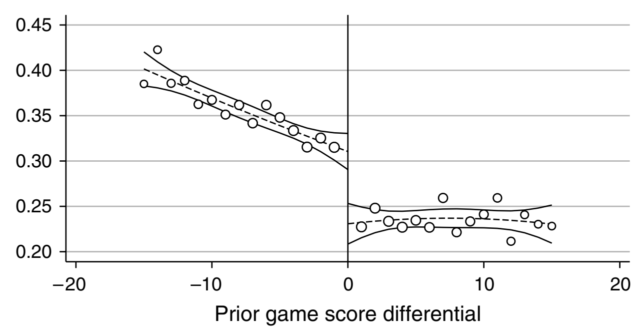
| Target | NBA | Simulated | Parameter | Value |
|---|---|---|---|---|
| Fraction of starter changes, given winning prior game | 23.3% | 23.42% | Exogenous strategy changes ($\tau$) | 0.233 |
| Fraction of starter changes, unconditional | 30.9% | 30.95% | Change in state ($\delta$) | 0.0074 |
| Fraction of wins | 50.0% | 49.98% | Bias ($\gamma$) | 1.56 |
| SD of score differential | 13.4 | 13.4% | SD of $P$ ($\sigma$) | 13.2 |
| Average of score differential | 0 | −0.01 | Average high performance ($h$) | 1.15 |
| Normalization | — | — | Average low performance ($l$) | −4.35 |

**Figure A.1 Histogram of Simulated Priors**




*Note.* The figure provides the distribution of priors from simulating $\rho_{b,t}$ over a sequence of 46,000 games, using biased updating between each game.

**Figure A.2 Probability Starters Are Changed as a Function of Prior Game Score Differential in Simulated Data**




*Notes.* The circles represent the mean outcome for each score differential value. The size of each circle is proportional to the number of games represented. The dashed line represents the predicted outcome, which was estimated with a second-order polynomial fully interacted with an indicator variable for whether the score differential was greater than or less than zero. The solid lines represent the 95% confidence intervals.

simulation, even when parameters have not changed. Thus, analysis on smaller samples of one team, one coach, or one team-season will be too unstable for the calibration process to converge. In any case, the NBA is, for the most part, competitively balanced across teams; hence, we anticipated that the targets would not be all that different among teams.

The calibrated parameters indicate that this is a highly competitive environment. Note that using the second-best strategy reduces average performance by only 5.5 points relative to the best strategy—although either strategy is highly noisy, with a standard deviation of 13.2 points. The unobserved shifts in environment ($\delta$) are rather infrequent but are still important to prevent beliefs from becoming too entrenched to respond to new information in the realized performance.

The bias parameter is calibrated to $\gamma = 1.56$, meaning that coaches overweight the outcome that occurred by 56%. This significantly distorts the posterior beliefs: after a win, the posterior is eight percentage points stronger under biased updating than it would be under Bayesian updating.[18] The distortion is also asymmetric: after a loss, the posterior is 11 percentage points weaker than under Bayesian updating. With any prior between 40.3% and 62.1%, the biased coach changes his strategy if and only if his team loses; for a Bayesian coach, this occurs only if the prior exactly equals 51.3%.

Ultimately, we are most concerned about how bias affects the frequency of strategy changes and the likelihood of winning. To determine how an unbiased coach would fare (holding all else equal, including the biases of other teams), we set $\gamma = 1$ and repeat the simulation. This differs from the procedure above only in that $\rho_{b,t}$ becomes the same as $\rho_{u,t}$—the coach's prior matches the true probability. To make the biased and Bayesian simulations comparable on a game-by-game basis, we use the same 46,000 realizations of the random variables $c_t$, $z_t$, $P_{t|h}$, and $P_{t|l}$. This exposes the Bayesian coach to the same luck as the biased coach—for instance, when both coaches select the right strategy ($c_t < \rho_{u,t}$), they will obtain the same performance, $P_{t|h}$. Thus, if the Bayesian coach performs better on average, it is only because he selects the right strategy more frequently (that is, the true probability of having the right strategy, $\rho_{u,t}$, is higher than for a biased coach).

In terms of strategy revision, the effect is remarkable: the starters are changed after only 26.1% of the games, rather than after 30.9%. Since strategy changes occur exogenously after 23.3% of games, biased updating increases elective strategy changes by nearly triple, from 2.8% to 7.6%. Note that this was not obvious a priori, since they change less frequently after winning and more frequently after losing, but the calibrated model reveals that the net effect creates more change than what is optimal. The asymmetry in bias is partly to blame for this, in that losing reduces the posterior to a greater extent than winning increases it.

The Bayesian coach is also more likely to use the correct strategy, which would reverse the outcome in 4.8% of the games. Of course, using the better strategy does not guarantee victory: in 2% of games, the biased coach chose the inferior strategy and still won because of good luck, yet the Bayesian coach chose the best strategy and lost ($P_{t|h} < 0 < P_{t|l}$). But more often (in 2.8% of games), the inferior strategy causes the biased coach to lose, whereas the superior strategy leads the Bayesian coach to win ($P_{t|h} > 0 > P_{t|l}$). The net effect is that, absent outcome bias, the coach would win 50.7% of the games instead of 50.0%.

Although this does not yield dramatic increases in wins, one should bear in mind that this is only one strategic choice in a highly noisy environment: the random events in a game create a standard deviation of 13.2 points, whereas the choice of starters can raise the average score by only 5.5 points. For instance, even if the coach always used the inferior strategy, he would win 37.3% of the games; using the optimal strategy every time, he would win 53.6% of the games. Thus, only 16.3% of the games can be affected by the decision-making process; among those, $0.7/16.3 = 4.3\%$ are lost because of outcome bias.

[18] This overconfidence, $\check{\rho} - \hat{\rho}$, is exactly 8% with prior $\rho = 0.65$ and performance $P = 8$. It grows larger when the prior is low or when performance is lower, but the change is fairly small. The same applies when computing underconfidence after a loss.

## References

Arcidiacono P, Kinsler J, Price J (2013) Productivity spillovers in team production: Evidence from professional basketball. Working paper, Duke University, Durham, NC.

Baron J, Hershey JC (1988) Outcome bias in decision evaluation. *J. Personality Soc. Psych.* 54(4):569–579.

Berger J, Pope D (2011) Can losing lead to winning? *Management Sci.* 57(5):817–827.

Bertrand M, Mullainathan S (2001) Are CEOs rewarded for luck? The ones without principals are. *Quart. J. Econom.* 116(3):901–932.

Camerer C, Loewenstein G, Weber M (1989) The curse of knowledge in economic settings: An experimental analysis. *J. Political Econom.* 97(5):1232–1254.

Card D, Dahl GB (2011) Family violence and football: The effect of unexpected emotional cues on violent behavior. *Quart. J. Econom.* 126(1):103–143.

Chiappori P-A, Levitt S, Groseclose T (2002) Testing mixed-strategy equilibria when players are heterogeneous: The case of penalty kicks in soccer. *Amer. Econom. Rev.* 92(4):1138–1151.

Denrell J, Liu C, Maslach D (2013) Bad luck or bad management: How uncertainty about others' inferences moderate responses to failures. Working paper, University of Warwick, Coventry, UK.

Dobbs IM (1991) A Bayesian approach to decision-making under ambiguity. *Economica* 58(232):417–440.

Earl PE (1990) Economics and psychology: A survey. *Econom. J.* 100(402):718–755.

Fischhoff B (1975) Hindsight ≠ foresight: The effect of outcome knowledge on judgment under uncertainty. *J. Experiment. Psych.: Human Perception Performance* 1(3):288–299.

Gino F, Moore DA, Bazerman MH (2008) No harm, no foul: The outcome bias in ethical judgments. HBS Working Paper 08-080, Harvard Business School, Boston.

Gino F, Shu LL, Bazerman MH (2010) Nameless + harmless = blameless: When seemingly irrelevant factors influence judgment of (un)ethical behavior. *Organ. Behav. Human Decision Processes* 111(2):93–101.

Gray PK, Gray SF (1997) Testing market efficiency: Evidence from the NFL sports betting market. *J. Finance* 52(4):1725–1737.

Gray R, Beilock SL, Carr TH (2007) "As soon as the bat met the ball, I knew it was gone": Outcome predication, hindsight bias, and the representation and control of action in expert and novice baseball players. *Psychonomic Bull. Rev.* 14(4):669–675.

Hawkins SA, Hastie R (1990) Hindsight: Biased judgments of past events after the outcomes are known. *Psych. Bull.* 107(3):311–327.

Jones SK, Yurak TJ, Frisch D (1997) The effect of outcome information on the evaluation and recall of individuals' own decisions. *Organ. Behav. Human Decision Processes* 71(1):95–120.

LaBine SJ, LaBine G (1996) Determinations of negligence and the hindsight bias. *Law Human Behav.* 20(5):501–516.

Lipshitz R (1989) "Either a medal or a corporal": The effects of success and failure on the evaluation of decision making and decision makers. *Organ. Behav. Human Decision Processes* 44(3):380–395.

Marshall GW, Mowen JC (1993) An experimental investigation of the outcome bias in salesperson performance evaluations. *J. Personal Selling Sales Management* 13(3):31–47.

McCrary J (2008) Manipulation of the running variable in the regression discontinuity design: A density test. *J. Econometrics* 142(2):698–714.

Offerman T, Sonnemans J (2004) What's causing overreaction? An experimental investigation of recency and the hot-hand effect. *Scand. J. Econom.* 106(3):533–553.

Pope DG, Schweitzer ME (2011) Is Tiger Woods loss averse? Persistent bias in the face of experience, competition, and high stakes. *Amer. Econom. Rev.* 101(1):129–157.

Price J, Wolfers J (2010) Racial discrimination among NBA referees. *Quart. J. Econom.* 124(4):1859–1887.

Rabin M (1998) Psychology and economics. *J. Econom. Literature* 36(1):11–46.

Ratner RK, Herbst KC (2005) When good decisions have bad outcomes: The impact of affect on switching behavior. *Organ. Behav. Human Decision Processes* 96:23–37.

Romer D (2006) Do firms maximize? Evidence from professional football. *J. Political Econom.* 114(2):340–365.

Tinsley CH, Dillon RL, Cronin MA (2012) How near-miss events amplify or attenuate risky decision making. *Management Sci.* 58(9):1596–1613.

Wolfers J (2002) Are voters rational? Evidence from gubernatorial elections. Stanford GSB Working Paper 1730, Stanford Graduate School of Business, Stanford, CA.