

The Role of Surprise: Understanding Over- and Underreactions Using In-Play Soccer Betting

Darwin Choi, HKUST
Sam K. Hui, NYU Stern*

February 2012

Abstract

We test the relationship between over-(under)reaction to news and the degree of surprise. Investors/bettors generally rely too heavily on prior beliefs and underreact, but sufficiently surprising information may cause overreaction due to the salience of the news. We use data from in-play soccer betting, where the arrival of information (goals) is well defined and the degree of surprise can be quantified. We find that bettors underreact to most goals, but overreact to highly surprising goals scored by “underdogs.” Our analysis suggests that over- and underreactions are subsequently corrected, and a profitable betting strategy could be formed based on the biases.

Keywords: Overreaction, Underreaction, Surprise, Betting Market, Anchoring, Salient Information, Correction of Biases.

JEL Classification: D8, G13, G14.

*Author Contact Information: Darwin Choi, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, dchoi@ust.hk. Sam K. Hui, New York University Leonard N. Stern School of Business, Tisch Hall 910, 40 West 4th Street, New York, NY 10012, khui@stern.nyu.edu. We thank Byoung Kang, Mujtaba Mian, Abhiroop Mukherjee, Mark Seasholes, and seminar participants at Hong Kong Polytechnic University for helpful comments. We acknowledge the General Research Fund of the Research Grants Council of Hong Kong (Project Number: 641011) for financial support. We also thank Chris Pang and Kaushik Ramakrishnan for excellent research assistance. All errors are our own.

1 Introduction

Most finance theories suggest that security prices are affected by how investors react to new information. However, whether investors over- or underreact to a particular one-time event is often a subject of debate. While the psychology literature suggests that people tend to overreact to strong signals in experiments, empirical evidence in financial markets is very limited because it is difficult to find an objective way to measure the strength of an information shock.¹ Taking a more extreme view, Fama (1998) argues that over- and underreactions are split randomly with no systematic patterns.

In this paper, we hypothesize that over- and underreactions to news depend on how “surprising” the news is. Our argument is built upon two well-known cognitive biases: anchoring to a reference point (Tversky and Kahneman, 1974; Kahneman, Slovic, and Tversky, 1982) and overweighting strong or salient information (Griffin and Tversky, 1992; Klibanoff, Lamont, and Wizman, 1998). A natural measure of salience is the degree of surprise: Itti and Baldi (2009) measure participants’ visual attention and show that surprising shocks are more salient and hence receive more attention. We therefore conjecture that, when reacting to new information that is in line with expectation or only moderately surprising, investors tend to anchor to their prior beliefs and hence underreact. However, new information that is very surprising may cause investors to overreact, as they overweight the information in their judgment due to the salience of the shock.

To test our hypotheses, we turn to data from an in-play soccer betting market.² In an

¹Klibanoff, Lamont, and Wizman (1998) use the appearance of news on the front page of *The New York Times* to proxy for salient information; however, it is a binary variable and does not quantify how salience the news is. Some papers in the finance literature classify over- and underreactions in other situations. For example, Barberis, Shleifer, and Vishny (1998) claim that investors generally underreact to earnings news but overreact to consistent patterns of earnings news; Daniel, Hirshleifer, and Subrahmanyam (1998) suggest that investors overreact to private signals and underreact to public signals. In our paper, we focus on the reaction to a one-time public event.

²We focus on the soccer betting market rather than other sports for two main reasons: First, goals in soccer arrive relatively infrequently and are hence important shocks. Second, the in-play soccer betting market is very active and typically has higher transaction volume compared to other sports. The total in-play betting volume in our sample of 2,017 international soccer matches over the period 2006–2011 amounts

in-play betting market, participants place bets while a match is still under way, and thus can actively react to new information (i.e., a goal) as it appears. While the sports betting market shares many structural similarities with financial markets (Avery and Chevalier, 1999), it offers several key advantages that allow us to study the role of signal strength in driving over- and underreactions as “cleanly” as possible. First, in soccer betting, the arrival of new information is apparent (goals) and its impact on financial payoffs can be directly benchmarked to the actual match outcomes. In contrast, “clean” empirical evidence of over- and underreactions is generally difficult to find in financial markets, because information in equity or derivatives markets may not be accessible to everyone at the same time and is not easy to separate from other market-wide news. Second, in the context of a soccer match, it is straightforward to define and measure how “surprising” a goal is by comparing the relative strength of the two teams: goals scored by underdogs are more surprising than goals that are scored by favorites. Third, goals in soccer are exogenous shocks to the in-play betting market; in financial markets, company information release may depend on managers’ perceptions of the firm being over- or undervalued.

Our data are comprised of second-by-second transaction records in 2,017 international soccer matches obtained from Betfair, the largest internet betting exchange in the world. We operationalize how “surprising” a goal is by the odds-implied probability of the non-scoring team minus that of the scoring team, measured before the goal.³ We document over- and underreactions to the first goal of the match using a sequence of logistic regressions of actual match outcomes (whether the scoring team wins) on odds-implied probabilities that the scoring team wins, as well as develop and estimate a formal Bayesian model that describes how these biases change over time and vary with betting volume. We find that, consistent with our hypotheses, within two minutes after the goal, there is evidence of systematic

to £2.99 billion (US\$4.86 billion).

³There are three possible outcomes for each match (Team 1 Win, Draw, and Team 2 Win). If we add up the reciprocal of the prevailing decimal odds of the three outcomes, the sum is typically very close to 1. We calculate the odds-implied probability using $\frac{1}{\text{odds}}$ divided by this sum, so that the odds-implied probabilities sum up to 1.

underreaction to goals that are not surprising or only moderately surprising. Underreaction is moderated by the degree of surprise, and the results point to overreaction if the surprise factor is large enough (around 0.41–0.56). Further, these biased reactions attenuate over time and disappear at around six minutes after the shock, but are unrelated to betting volume.

Finally, we form profitable strategies by betting on the scoring team when underreaction is predicted, and against the scoring team if overreaction is predicted. In our sample such strategies earn a profit of 2.79% (p -value = 0.02) after commissions if the bets are placed at two minutes after the goal, 0.82% (p -value = 0.26) if six minutes. This finding not only confirms that the biased reactions are present in the in-play markets and are corrected subsequently, but also implies that the results are economically significant, particularly when the total volume of all matches in the [+2, +3 minutes] window betting in the direction of our strategies already exceeds £80 million.

Our paper makes three key contributions. First, we build a novel dataset to study reactions to information; to the best of our knowledge, our paper is one of the first to use the in-play betting market to identify over- and underreactions. Second, we empirically demonstrate that bettors react more strongly to surprising shocks; this extends findings in psychological experiments to real-life situations involving financial stakes and generates predictions for reactions to information events in financial markets. Finally, although we could not identify traders in the data, the amount of profits is an estimate of the total wealth transfer from irrational bettors to arbitrageurs; it provides a way to measure the cost of investor biases.

The remainder of this paper is structured as follows. Section 2 briefly reviews the related literature on over- and underreactions in financial markets and previous research on sports betting markets. Section 3 outlines our major hypotheses. The data on Betfair bets and soccer matches are discussed in Section 4. Section 5 provides empirical evidence of over- and underreactions, as well as constructs a trading strategy based on the biases. Section 6

discusses the implications for financial markets.

2 Related Literature

2.1 Over- and Underreactions in Financial Markets

We briefly review previous literature that studies over- and underreactions to new information in financial markets. Several papers have examined underreaction to information. For example, in Hong and Stein's (1999) model, investors are only able to process some subset of public information; as a result, the slow diffusion of information among bounded-rational agents causes short-run underreaction. Taking an empirical perspective, Klibanoff, Lamont, and Wizman (1998) document that close-end country fund prices generally underreact to changes in fundamentals, measured by the Net Asset Value (NAV); but the degree of underreaction is lower if there is salient news (e.g., if it appears on the front page of *The New York Times*). Grinblatt and Han (2005) propose and test an explanation for underreaction to information based on the tendency of investors to hold losers and sell winners, i.e., the disposition effect. On the other hand, while Daniel, Hirshleifer, and Subrahmanyam (1998) also show that stock investors generally underreact to public information, they find that investors tend to overreact to private information. They explain this phenomenon by investors' overconfidence: investors overestimate the precision of their private signals and rely too much on those signals.

A related stream of literature looks at over- and underreactions to corporate events. For example, Barberis, Shleifer, and Vishny (1998) develop a framework to produce underreaction to earnings news due to conservatism, as well as overreaction to a series of earnings news based on representativeness.⁴ Frazzini (2006) shows that the disposition effect contributes

⁴This is supported by Poteshman (2001), who shows daily underreaction and multiple-day overreaction of option prices to shifts in volatility.

to underreaction to earnings news. Ramalingegowda, Shu, and Yeung (2011) argue that investors underreact to firm’s earnings but overreact to blockholder’s earnings. They attribute their results to moderated confidence (Bloomfield, Libby, and Nelson, 2000), where precise signals are underweighted and imprecise signals are overweighted. Michaely, Thaler, and Womack (1995) find underreaction to dividend initiations and omissions. For more examples of corporate finance event studies that interpret their findings as over- or underreactions, see Thaler (1993, 2005).

The above literature review identifies several limitations of the previous research on over- and underreactions, and motivates the current study. First, though previous research has examined several factors that affect over- and underreactions, very few papers have explicitly studied the strength of the shock in driving over- and underreactions. Second, we note that “clean” empirical evidence is generally difficult to find in financial markets. Some phenomena can be viewed as evidence of opposite claims. One example is momentum in stock prices (Jegadeesh and Titman, 1993). While Barberis, Shleifer, and Vishny (1998) and Hong and Stein (1999) argue that momentum is caused by subsequent correction to underreaction, De Long et al. (1990) and Daniel, Hirshleifer, and Subrahmanyam (1998) attribute the same phenomenon to overreaction. Thus, the current research contributes to the existing literature by studying the role of surprise in over- and underreactions using data from the in-play soccer betting market.⁵ Next, we present a review of the previous literature that studies sports betting markets.

⁵Participants and financial stakes in betting markets are different from those in the stock market. Similar concerns are raised by Levitt and List (2007a, b) about the applicability of experimental settings to financial markets. In particular, in betting markets expected returns are usually negative and participants’ risk preferences may differ from stock market participants’. However, the goal of our paper is to document biased reactions to new information, which we believe are based on well-known cognitive biases and should apply to typical investors.

2.2 Sports Betting Markets

Several researchers have examined the efficiency of sports betting markets. Early works focus on studying the efficiency of “pre-match” betting markets, where bettors are only allowed to place their bets before a match begins and payoffs are made after the match ends.

Gandar, Zuber, O’Brien, and Russo (1988) do not find statistical evidence to reject market efficiency using National Football League (NFL) betting data. Golec and Tamarkin (1991) develop more powerful tests and find a bias against underdogs in the NFL betting market. Gray and Gray (1997) and Avery and Chevalier (1999) formulate profitable betting strategies by exploiting this and other biases. Vergin (2001) shows that NFL bettors tend to overweigh positive performance over recent games or the previous season. On the other hand, a survey article by Thaler and Ziemba (1988) notes that in racehorse betting markets there is a bias against favorites, which is often called the favorite–longshot bias. Using college football markets, Durham, Hertz, and Martin (2005) find limited evidence that investors are influenced by trends and patterns as suggested by Barberis, Shleifer, and Vishny’s (1998) model, while Durham and Santhanakrishnan (2008) show that bettors overreact to average points in excess of the spread (as a proxy for strength) and underreact to streak against the spread (as a proxy for weight). Another type of markets that is similar to sports betting is prediction markets, where payoffs depend on unknown future events such as presidential elections. These markets are analyzed by, for example, Wolfers and Zitzewitz (2004, 2006) and Arrow et al. (2008).

More recent research has started to look at in-play sport betting markets, where bettors are able to place bets while the match is still under way. Betting exchanges now allow researchers to study the evolution of real-time betting odds during the match. A pioneering work by Gil and Levitt (2007) collect data on fifty Soccer World Cup matches in 2002 from a small internet betting exchange (intrade.com). They find that odds that a certain team wins tend to continue drifting after ten to fifteen minutes after a goal, which they

interpret as violations of market efficiency. This claim is disputed by Croxson and Reade (2011), who argue that even if odds are efficient, in-play odds should continue to drift after a goal is scored. To see this, consider a match between team A and team B, and suppose that team A scores the first goal at the 75th minute and hence lead by 1-0. Under the assumption of efficient odds, the odds that team A wins should continue to increase till the end of the match if no goals are scored during the period, because team A will hold on to its one-goal margin. Expanding upon this argument, Croxson and Reade (2011) use data from Betfair and examine the real-time odds when scores change, with the goal of testing market efficiency. They assemble a dataset of over 1,200 soccer matches, along with the corresponding scoring data. An important insight drawn from their research is that in order to accurately assess market efficiency with betting odds data, one must carefully incorporate a model of how “efficient” prices must drift as time passes. Building upon this insight, we develop and estimate a formal Bayesian framework in Section 5.4 that incorporates a model of “efficient” prices, hence allowing us to assess the degree of over- and underreactions. Our result is different from Croxson and Reade’s (2011) in that we show over- and underreactions to first goals, as well as the subsequent correction; our goal is not to test the efficiency in this particular market, but to identify biased reactions using the unique setting.

3 Hypotheses

In this section, we develop our key hypotheses about how over- and underreactions to new information are driven by how “surprising” the information is, and discuss in detail how we will test our hypotheses using data from the soccer betting market. The framework we propose here is based on two cognitive biases, anchoring and salience. These psychological beliefs are particularly relevant because when reacting to new information (e.g., macroeconomic events or company news in financial markets, goals in soccer betting markets), market participants have to make a split-second decision and are thus more likely to rely on various

behavioral heuristics in making predictions and judgments under uncertainty.

First, we argue that in general, market participants are likely to underreact to new information due to anchoring (Tversky and Kahneman 1974), which refers to the insufficient adjustments from a pre-existing reference point. For instance, Wright and Anderson (1989) observe a strong anchoring effect when participants in an experiment are asked to judge whether the event probability is above or below 0.25 (low anchor) or 0.75 (high anchor), before making final assessments. The final probability judgment is greater in the high anchor case. Similarly, using a multi-stage experiment that represents a random walk with upward drift, Gneezy (1996) claims that subjects apply the stage-by-stage probability as an anchor and adjust insufficiently, leading to underestimation of the overall probability. Building upon the previous literature, we hypothesize that, on average, investors tend to underreact to new information.

Second, we further hypothesize that this general pattern of underreaction is moderated by how “surprising” the new information is. Previous literature in psychology and decision science finds that people tend to assign more “weight” to information that is salient; as a result, events that are distinctive, obvious, and prominent affect judgments disproportionately. For instance, Griffin and Tversky (1992) show that people overreact to the strength or extremeness of the evidence (e.g., the warmth of a recommendation letter) and underreact to weight (e.g., the credibility of the person who writes the letter). Further, by measuring participants’ visual attention, Itti and Baldi (2009) show that information that is more surprising draws more attention; Ranganath and Rainer (2003) review research in neuroscience and suggest how novel events cause neural responses that result in greater attention to and memory for those events. Thus, we hypothesize that the more surprising an information shock is, the lower the degree of underreaction, as under-adjustment due to anchoring is offset by the overweighting due to the salience of the new information. Taking this one step further, if the new information is extremely surprising, the overweighting of salience information will likely more than offset the effect of anchoring, resulting in overreaction instead.

We now turn to the context of the in-play soccer betting market, which will be used to test our proposed hypotheses. In soccer betting, the odds of the bets specify the payoffs bettors would receive if a certain outcome happens. For example, a bet on a certain team winning the match with decimal odds of 5 will receive \$5 for every \$1 of bet if the team wins (including the original \$1). The implied probability of this event from the odds is $\frac{1}{5}$ in a fair game with zero expected return. Therefore, betting odds reflect the aggregate market's beliefs of the winning probabilities of the teams, and can be used to gauge whether bettors over- or underreact to goals. More specifically, if bettors overreact to a goal, they rush to bet on the scoring team and bid down the odds, inflating the implied probability that the team wins. If they underreact to a goal, the odds of the scoring team will be too high.

Applying our proposed hypotheses to the soccer betting setting, we hypothesize that bettors' reactions to the first goal of a match depend on how surprising that goal is. Note that we study only the first goal of each match because the changes in winning probabilities should be more significant for these goals. In contrast, the fifth goal in a 5-0 game does not materially affect bettors' beliefs, and may not even attract their interest (since there are other markets such as betting on the full time score).

We operationalize "surprise" using the strength of the two teams: the favorite (underdog) is defined as the team having a higher (lower) winning probability shortly before the first goal of the match is scored. When the first goal of the match is "not surprising" (the scoring team is the favorite), anchoring will dominate the updating of the beliefs: the favorite is still the favorite and insufficient adjustments are made. However, underreaction is moderated by how surprising the goal is, and overreaction occurs when the first goal is very surprising (i.e., the scoring team is much weaker). This is because prior beliefs may need to be changed dramatically (it is unclear which team is the favorite after the goal) and, as hypothesized earlier, bettors tend to overweight the importance of salient shocks.

To sum up, we state our key hypothesis (in the context of soccer betting) as follows:

H1 (Over- and underreactions to the first goal): Whether bettors exhibit over- or underreactions depends on how surprising the first goal is. Bettors generally underreact to the first goal. However, the degree of underreaction decreases with surprise, with bettors overreacting to a highly surprising first goal.

Next, we propose two other hypotheses (H2, H3) on how the biases stated in H1 vary with time lag and betting volume. Anecdotal evidence suggests that professional bettors participate actively in the in-play market and have designed software to place bets.⁶ We therefore expect that the biases in H1 are short-lived as arbitrageurs may form strategies to profit from the over- and underreactions. Our objective is to formally test the correction and document the amount of time needed for it. Soccer matches last only 90 minutes (plus any injury time) and the outcome is realized at the end of the game, but we are interested in assessing to extent to which the correction happens earlier:

H2 (Time moderation): The bias found in *H1* attenuates over time after the first goal is scored.

Finally, we also examine if higher volume corresponds to lower bias. Although our data do not have traders' identity, one possible way to distinguish professional traders from small bettors is to study transaction volume. The intuition is that higher volume involves larger financial stakes and could be placed by more rational and/or informed traders. It is, however, also possible that these traders split their orders into smaller bets to hide their activity; in this case transaction volume cannot be used to identify bettors.⁷ Our last hypothesis

⁶See, for example, "The super punters who can't lose; A sure bet," *Herald Sun* (Melbourne, Australia), 10/30/2004. Some bettors place bets on Betfair using automated software ("bots").

⁷In the context of financial markets, Easley and O'Hara (1987) develop a model and show that informed traders prefer to trade larger amounts; empirically, Hasbrouck (1988, 1991) shows evidence that large trades contain more information; however, Barclay and Warner (1993) find medium-sized trades more informative, suggesting that informed traders break up their trades.

tests whether there is any systematic relationship between the biased reactions and betting volume:

H3 (Volume moderation): The bias found in *H1* decreases with higher betting volume.

The empirical results of these hypotheses are shown in Section 5. The next section describes our data and sample of soccer matches.

4 Data

We test our hypotheses using data from Betfair, an online sports betting exchange platform. Betfair was founded in 1999 and is currently the world's largest internet betting exchange. In addition to soccer, it also offers betting in many other sports, as well as activities such as horse racing and poker. On its website, Betfair claims to process more than seven million transactions a day, more than all European stock exchanges combined. The company is based in the U.K. and accepts online bets from different countries, but excluding the U.S. because of legal restrictions. Historical data are available for purchase from an authorized third-party vendor, Fracsoft.

We obtain from Fracsoft all available matches of major international and European soccer competitions, both at the club and national team levels, from the period August 2006 to March 2011. Specifically, our dataset is comprised of a total of 2,687 soccer matches in the following competitions: FIFA World Cup 2006 and 2010, five seasons of UEFA Champions League in Europe and Barclays Premier League in England (2006/07 to 2010/11), as well as one season of La Liga in Spain, Serie A in Italy, and Bundesliga in Germany (2010/11).⁸ Betfair operates several markets for each match, including Match Odds (the outcome in 90 game minutes: Team 1 Win, Draw, Team 2 Win), Half Time and Full Time Score, and

⁸Note that not every match is covered by Fracsoft. In an email correspondence with Fracsoft, they attribute the missing data to technical issues when recording the bets from Betfair. Nevertheless, most matches (over 70%) are available. We believe the selection is random and thus does not affect our results.

Total Number of Goals Over/Under 2.5, etc. Betfair allows participants to “back” (bet on) or “lay” (bet against) an outcome. As discussed earlier, a “back” bet on a certain team winning the match with decimal odds of 5 will receive \$5 for every \$1 of bet if the team wins (including the original \$1), while a corresponding “lay” bet will lose \$5 if the team wins and earn \$1 otherwise. As an exchange Betfair matches the orders, and therefore the matched bets are zero-sum games between the “back” and “lay” bettors. The exchange charges a commission of up to 5% on net winnings after the outcome is known, and the commissions are not reflected in the decimal odds.⁹

In our analysis, we focus on the matched bets in the “Match Odds” market. The data record the volume, decimal odds, timestamps (to the second), and outcomes betting in favor of or against for all the bets. All data within a second are aggregated by Fracsoft. Participants can place bets before the game starts as well as when the soccer match is underway. For almost all matches, Betfair suspends trading briefly (for around two minutes) and clears all unfilled “back” and “lay” orders when the game starts and when a “material event” has happened. Material events include goals, penalty kicks awarded, and red cards (which result in player’s dismissal). However, while we could observe the time when trading is suspended, the Betfair data do not record the actual material events leading to the suspension.

We therefore augment our dataset by collecting event information from ESPN Soccernet (soccernet.espn.go.com). The events are merged with the Betfair suspended trading time using the event time. To ensure that the merging procedure is as accurate as possible, we first obtain the starting time of the game from Betfair when the market just turns “in-play,” then estimate the event time, and compare that with Betfair’s suspended trading time. Note that the data are somewhat “noisy”: there are several cases where trading is not suspended despite the record of a material event, and some cases where trading is suspended without a corresponding material event. In all these cases, we have applied our best judgment to

⁹This structure makes Betfair different from other traditional betting platforms, where the odds are typically set by a bookmaker. Betfair matches orders from bettors and does not set the odds. Therefore, any biases we observe from Betfair odds should come from market participants rather than a bookmaker.

ensure that we get the correct event by studying game reports on ESPN Soccernet and check the change in odds before and after trading suspension. We only include matches that did suspend trading after the first goal in our sample.

Out of all matches, 2,160 of them have at least one goal scored and the first goal’s scoring time could be identified from the Betfair data. We exclude first goals that are scored from penalty kicks, because Betfair suspends trading when penalties are awarded, not when they are scored. This results in a final sample of 2,017 matches. Table 1 presents some summary statistics of our dataset. The matched in-play volume per match is around £1.5 million on average, with a minimum of £9,654 to a maximum of over £13.4 million. The first goal of the match is, on average, scored around 36 minutes after the match has started. We also calculate the betting volume in the window [+2, +6 minutes] after the first goal. The average is £152,523, or around 10% of the total in-play volume. 67% of the first goals are scored by the favorite, and 71% of the scoring teams eventually win the match. We also show the statistics of the odds-implied probabilities at 2, 3, and 6 minutes after the first goal (these probabilities are described in Section 5.1). As can be seen, there appears to be some initial evidence of general underreaction: the average odds-implied probabilities (0.69) are (slightly) lower than the actual proportion of wins (0.71).

5 Over- and Underreactions: Evidence and Implications

In this section, we formally test hypotheses H1–H3 that we proposed in Section 3 using several statistical approaches. In Section 5.1, we operationalize how “surprising” the first goal is. Section 5.2 presents some descriptive analyses exploring how the relationship between the odds-implied probability and actual match outcome is moderated by “surprise.” Using the match-level data, Section 5.3 estimates a sequence of logistic regressions that provide some

initial evidence on over- and underreactions. Taking a step forward, in Section 5.4 we develop and estimate a formal Bayesian model of second-by-second transaction data. Our model incorporates a parametric probability model of how (latent) efficient odds evolve (Croxson and Reade, 2011). The formal model allows us to assess the extent to which over- and underreactions are moderated by time lag (H2) and transactional volume (H3). Having established statistical evidences that support our hypotheses, we compute the economic significance of over- and underreactions in Section 5.5, and explore several alternative explanations of the observed phenomenon in Section 5.6.

5.1 Operationalization of the Surprise Metric

As discussed in Section 3, we hypothesize that the extent to which the anchoring or salience heuristic dominates depends on how “surprising” the first goal is. Here, we quantify the degree of surprise of the first goal using the odds data. We define s_i , the “surprise” metric, as the following:

$$s_i = \textit{Implied-Pr}_i(\text{non-scoring team wins}) - \textit{Implied-Pr}_i(\text{scoring team wins}), \quad (1)$$

where both odds-implied probabilities ($\textit{Implied-Pr}_i$, as defined below) are measured at one minute before the first goal of match i .¹⁰ A positive (negative) s_i means that the scoring team is relatively weaker (stronger) than the non-scoring team, and that the goal is surprising (expected).

¹⁰This definition captures the market’s perception of the relative strength of the two teams, right before a goal is scored. We note that, however, if the score remains at 0-0 till very late in the match, the odds-implied probability of draw will be high, and the difference in the odds-implied probabilities of the two teams tends to be small. As a robustness check we use odds-implied probabilities before the match starts, calculated using pre-match odds, which are unaffected by the time of the first goal. The results using this alternative surprise metric are similar and are not reported. Also, Betfair provides a market to bet on or against the “Next Goal” (Team 1 Scores, Team 2 Scores, or No Goal) and one could potentially calculate another surprise metric based on the odds in this market. But the matched betting volume in the Next Goal market is much lower than the Match Odds market, and we decide to focus on the latter.

The odds-implied probabilities are calculated from the decimal odds as follows:¹¹

$$\text{Implied-Pr}_i(\text{team 1 wins}) = \left[\frac{1}{\text{odds}_i(\text{team 1})} \right] / c_i, \quad (2)$$

where

$$c_i = \frac{1}{\text{odds}_i(\text{team 1})} + \frac{1}{\text{odds}_i(\text{draw})} + \frac{1}{\text{odds}_i(\text{team 2})}, \quad (3)$$

again all measured at one minute before the first goal of match i (i.e., the odds are the last transacted odds one minute before the goal). The normalizing constants c_i 's are usually very close to one, and are included to ensure that the implied probabilities of all outcomes sum up to one.

In our sample, s_i ranges from -0.92 to 0.88 , with a mean of -0.14 and a median of -0.14 (a negative mean/median is expected because the favorite is more likely to score the first goal). A histogram of s_i across 2,017 matches is shown in Figure 1.

For illustration purposes, we show how the odds-implied probability of the scoring team reacts to the first goal using two matches in our sample. The top graph in Figure 2 shows a “surprising” goal ($s_i = 0.75$), scored by New Zealand (the underdog) against Italy in the FIFA World Cup 2010; while the bottom graph shows an “expected” goal ($s_i = -0.72$), scored by Arsenal (the favorite) against Watford in the Barclays Premier League 2006/07. The probability goes up in both cases, reflecting the increased likelihood that the scoring team will win given that it is now leading by one goal to nil. Notice that there is a lack of transaction activities right after the goal because trading is suspended for around 1–2 minutes after the goal is scored, and the market is reopened shortly afterwards. Note also that by looking at the graphs alone, we cannot infer whether the odds of a particular match at a particular time show over- or underreactions, because we do not know the true

¹¹The definition of implied probabilities is similar to Asch, Malkiel, and Quandt (1984), who study bettors' subjective probabilities in racetrack betting. While they include the “take” that is subtracted from the total betting pool in their calculations, we exclude Betfair's commissions because each bettor pays a fixed commission for all bets in a given match (and so including and excluding it give the same probabilities).

(unobserved) winning probability of the scoring team. Thus, statistical evidence of over- and underreactions can only be identified by comparing the odds across a set of matches to the corresponding set of outcomes, a strategy that we pursue next in the following subsections.

5.2 Descriptive Analyses of Over- and Underreactions

To assess Hypothesis H1, we begin by conducting some descriptive analyses that explore the relationship between actual match outcomes and odds-implied probabilities, for different values of s_i . Recall that after the first goal, the betting market is suspended for around two minutes before it reopens. Thus, to assess how bettors update their beliefs after the first goal, we compare the in-play odds two minutes after the first goal is scored (the last transacted odds before the two minute-mark) versus the actual match outcomes.

Specifically, we compute the following sets of variables. We first calculate the implied probability p_i that the scoring team wins ($= \text{Implied-Pr}_i(\text{scoring team wins})$) using odds two minutes after the first goal, using equations (2) and (3).¹² We let y_i be an indicator variable which takes a value of 1 if the scoring team in match i wins the match, and 0 otherwise. If bettors' reaction to the first goal is efficient, p_i should be equal to the conditional probability that $[y_i = 1]$ given all information available at that time. Thus, we compare the expected count that the scoring team wins from these implied probabilities ($\sum_i p_i$) versus the actual count that the scoring team wins ($\sum_i y_i$), and use the difference as a statistical measure of over- or underreactions. If there is over-(under)reaction to the first goal, the expected count should be larger (smaller) than the actual count.

To explore how over- and underreactions are moderated by surprise, we divide the data

¹²While we focus on the scoring team, the hypothesis is unchanged if we look at the draw odds and the non-scoring team's odds. When a goal is scored by Team 1, Team 1 odds should decrease and Draw and/or Team 2 Win odds should increase. Our hypothesis defines underreaction as insufficient adjustments in the scoring team odds (i.e., the odds are not low enough). Since the sum of $\frac{1}{\text{odds}}$ over all outcomes is very close to 1, this is equivalent to insufficient adjustments in the draw and/or the non-scoring teams odds (i.e., the odds are not high enough). The same argument applies to overreaction.

into two sets: “expected” goal, which corresponds to the condition ($s_i < -c$), and “surprise” goal, which corresponds to the condition ($s_i > c$). We vary the threshold value c from 0.0 to 0.6, and in each case compare the total number of times that the scoring team wins with its expected value under the assumption that reactions to first goals are efficient; the results are shown in Table 2. A sequence of p -values, which provides some indication whether the observed data are different from their expected values, is generated using the following procedure:

1. As discussed, the null hypothesis H_0 is that the in-play odds at the two-minute mark (p_i) accurately reflects the probability of the match outcome (y_i), i.e., $H_0 : Pr(y_i = 1) = p_i$.
2. The test statistic employed here is the difference between the expected number of times that scoring team wins and the actual number of wins, i.e., $d = \sum_i p_i - \sum_i y_i$.
3. The null distribution of test statistic d is generated by a Monte Carlo simulation under the null hypothesis that y_i is *i.i.d.* Bernoulli with probability p_i .

As can be seen on the left column of Table 2, the expected number of wins by scoring teams is generally smaller than the actual observed number when the first goal is “expected” ($s_i < -c$), indicating that bettors tend to underreact to an “expected” shock. This effect is statistically significant at $p = 0.05$ level for most values of the threshold c (except when $c = 0.4$, which is significant at the $p = 0.10$ level). However, by contrasting this with the right column of Table 2, we find that this significant underreaction disappears when the first goal is a “surprise” ($s_i > c$). Interestingly, the deviations of actual vs. expected number of wins by the scoring teams go in the opposite direction when the goal is very “surprising” ($s_i > 0.4$): the expected number of wins by the scoring team is generally larger than the actual observed number. Note, however, that none of the p -values for the “surprise” condition is statistically significant, presumably because of the small sample size which results in the statistical test having low power.

Therefore, our descriptive analyses above provide some initial evidence that is consistent with H1: bettors tend to underreact to expected shock ($s_i < 0.0$), but overreact to surprising shock (when $s_i > 0.4$). To explore H1 in more detail, and to study the moderating factors of time lag and transaction volume proposed in H2 and H3 (respectively), we now follow up with a sequence of more elaborate statistical models and some additional parametric assumptions.

5.3 A Sequence of Logistic Regressions

We now estimate the magnitude of the bias in H1 as a function of s_i , and study how the bias in H1 is moderated by time lag, as hypothesized in H2. Towards that end, we use equations (2) and (3) to calculate the odds-implied probability for each match at $t = 2, 3, \dots, 15$ minutes after the first goal (we use the last transacted odds before each minute-mark).¹³ We then estimate the following sequence of logistic regressions:

$$y_i \sim \text{Bernoulli}(\pi_{it}) \tag{4a}$$

$$\text{logit}(\pi_{it}) = \text{logit}(p_{it}) + \alpha_t + s_i\beta_t, \tag{4b}$$

where, as before, p_{it} denotes the odds-implied probability (that scoring team will win) at the t -th minute after the first goal; y_i is an indicator variable that takes the value of 1 if the scoring team wins the match, and 0 otherwise; π_{it} denotes the “true” conditional probability that $y_i = 1$ given all the information available at time t . The term $(\alpha_t + s_i\beta_t)$ represents the direction and magnitude of the bias as a function of s_i (surprise). Clearly, $\alpha_t + s_i\beta_t > 0$ means that the true probability π_{it} is larger than the odd-implied probability p_{it} ; it means

¹³Of course, other goals and material events can occur in this timeframe. We do not filter our sample based on whether other events occur or not, because the reactions after the first goal should also take into account the expectation of future events. Also note that the sample size gets smaller for logistic regressions of longer time lags. For example, the one with time lag = 15 minutes excludes all matches with the first goal scored at less than 15 minutes before the match ends. Section 5.4 develops a Bayesian model that does not require long time lags; as will be discussed, the results are similar to those presented here.

that bettors do not make sufficient adjustment to the winning probability of the scoring team (i.e., underreaction). Conversely, $\alpha_t + s_i\beta_t < 0$ indicates overreaction. The models in (4) are estimated using the package *glm()* in R.

For each value of t , we perform a likelihood ratio test (Nelder and Wedderburn, 1972) to compare fit of the model described by equation (4) against the “efficient” model, which refers to the joint null hypothesis that $\{\alpha = \beta = 0\}$. Under the “efficient” model, $\pi_{it} \equiv p_{it}$, and hence the odds-implied probability accurately reflects the conditional probability of the match outcome. Thus, a small p -value provides statistical evidence for over- or underreactions, depending on the sign of $\alpha_t + s_i\beta_t$. Further, if the systematic bias in H1 is attenuated by time lag, we should find that as time goes on, the p -value should increase and at some point becomes insignificant.

This is exactly borne out by the results shown in Table 3. First, we find that the best fitting model for the in-play odds at $t = 2$ (i.e., two minutes after the first goal, when the betting market just re-opens) is:

$$\text{logit}(\pi_{it}) = \text{logit}(p_{it}) + 0.157 - 0.385s_i, \quad (5)$$

with a p -value of 0.001, showing that this model fits significantly better compared to the “efficient” model. This rejects the null hypothesis of efficient in-play odds, in favor of the alternative hypothesis (H1) that in-play odds are subject to systematic bias, described by the term $\alpha_t + s_i\beta_t$. As shown in Figure 1 and in Section 4, the surprise metric s_i ranges from -0.92 to 0.88 , so $\alpha_t + s_i\beta_t$ (for $t = 2$) ranges from 0.51 to -0.18 .¹⁴ This indicates that whether underreaction ($\alpha_t + s_i\beta_t > 0$) or overreaction ($\alpha_t + s_i\beta_t < 0$) is observed depends on the value of s_i . Specifically, when $s_i > 0.41$, we observe overreaction. This corroborates

¹⁴The over- and underreactions are quoted in logit scale; thus, the magnitude of the bias terms depends on where the true probability is on the logit scale. For instance, if we assume that the true probability is 0.5, an underreaction of 0.51 in logit scale would translate to an underestimation of 0.125 in the probability (from 0.5 to 0.375), while an overreaction of -0.18 would translate to an overestimation of 0.045 in the probability (from 0.5 to 0.545).

with the initial findings in the descriptive analysis presented in Section 5.2: when the first goal is highly surprising, bettors tend to overreact and overestimate the likelihood that the underdog will win.

Second, we find that the bias in H1 is attenuated over time, which is consistent with H2. We see that the p -value in the last column (which measures the model with bias vs. the efficient model) increases with time. The p -value is still below 0.05 by the fourth minute, but becomes marginally significant ($p = 0.07$) at the fifth minute. The p -value becomes insignificant ($p = 0.11$) after six minutes, indicating that the odds have stabilized and that the “efficient” hypothesis can no longer be rejected. This pattern holds also for the individual p -values that test $\alpha = 0$ and $\beta = 0$; both p -values increase over time and are significant only for the first 3–4 minutes after the first goal. This suggests that the systematic bias in H1 is corrected in around 5 minutes.

To further explore the robustness of the parametric assumptions in equation (4), we also conduct a Hosmer-Lemeshow (HL) test (Hosmer and Lemeshow, 2000) for the in-play odds 2 to 15 minutes after the first goal is scored. The HL test, which compares a probability assessment versus actual outcomes without making any parametric assumptions, is a commonly used goodness of fit test in applied statistics (Hosmer and Lemeshow, 2000). The test statistic follows a χ^2_{k-2} distribution, where k is the number of “bins” used to group the data (chosen to be $k = 20$ here). As can be seen in Table 4, the in-play odds two to four minutes after a goal is scored rejects the Hosmer-Lemeshow test ($p < 0.1$), indicating that in-play odds are not efficient; the p -value becomes insignificant after the fifth minute, which corroborates our findings with the sequence of logistic regressions above.

5.4 A Formal Bayesian Model of In-Play Odds

In Section 5.3, we show that the bias described in H1 (i.e., underreaction to expected and moderately surprising shocks and overreaction to very surprising shocks) attenuates over

time, and gradually disappears by the fifth and sixth minute after the first goal, by looking at the data at an aggregate level. We did not, however, provide an estimate of the *rate* at which the bias is reduced over time. Further, the logistic regression approach does not control of the occurrence of other material events (e.g., red cards, another goal, penalty kick) after the first goal. In this section, we now analyze the data at a second-by-second level, by developing a formal Bayesian model to estimate the extent to which the bias decreases over time, and in addition explore the relationship between over- and underreaction biases and transactional volume.¹⁵

We now set up a formal Bayesian model that allows us to incorporate relevant domain knowledge, by specifying prior distributions on model parameters, to estimate the patterns of over- and underreactions in in-play odds. As before, let i index matches ($i = 1, \dots, I$), and y_i denotes the match outcome ($y_i = 1$ if the scoring team wins the match, and 0 otherwise). Because we now analyze second-by-second data, we will change our notation of time slightly, by using t to index time in seconds rather than minutes. We let t_i^0 denote the running time (time since the match begins), in seconds, in match i when the first goal is scored; t_i^r denote the running time in match i when the market re-opens after the suspension following the first goal. As discussed earlier, t_i^r is about 60–120 seconds (1–2 minutes) after t_i^0 . Further, let V_{it} denote the (mean-centered) log-transaction volume of the matched bet at time t of match i .

Let $\tilde{p}_{it} = Pr(y_i = 1 \mid F_{it})$ be the true (yet unobserved) conditional probability that the scoring team in match i will win the match; given all the information (F_{it}) available up to

¹⁵This is somewhat different from using high-frequency financial data to study issues in asset pricing. Our data are aggregated by Fracsoft at the second-by-second level, while market microstructure studies typically analyze transactions data, which are irregularly spaced. Because of the short timeframe and high trading activity in the betting market, we are assuming that our second-by-second analysis is close to day-by-day analysis using stock price data. Therefore, we are not considering the special econometric concerns (see, e.g., Engle, 2000) associated with high-frequency stock price data.

running time t . By the definition of $\tilde{p}_{it_i^r}$, we have:

$$y_i \sim \text{Bernoulli}(\tilde{p}_{it_i^r}). \quad (6)$$

Next, we specify a prior distribution for $\tilde{p}_{it_i^r}$ by modeling the scoring process of a soccer match using an independent bivariate Poisson process (Maher, 1982). We specify the following model:

$$\text{logit}(\tilde{p}_{it}) = \text{logit}(\tilde{p}_{i,t-1}) + g_{it} + \varepsilon_{it} \quad (t > t_i^r) \quad (7)$$

$$\varepsilon_{it} \sim N(0, \sigma_{it}^2), \quad \text{where } \sigma_{it}^2 = \begin{cases} \sigma^2 & \text{if } E_{it} = 0 \\ \tau^2 & \text{if } E_{it} = 1 \end{cases} \quad (\sigma^2 < \tau^2), \quad (8)$$

where g_{it} denotes the “efficient drift” in $\text{logit}(\tilde{p}_{it})$, under the assumption that the scoring process in a soccer match follows a bivariate Poisson process, given the events that occur between time $t - 1$ and t . Derivation and computation of the bivariate Poisson process and g_{it} are described in more detail in Technical Appendix I. E_{it} is an indicator variable which takes a value of 1 if another “material event” (e.g., another goal, penalty awarded, and red card) occurs at time t . The specification in equation (8) reflects that in the absence of another “material event” at time t , the absolute magnitude of the modeling error ε_{it} should be smaller than if another “material event” has occurred ($\sigma^2 < \tau^2$).

Next, we model the observed in-play odds-implied probability as the sum of the true conditional probability, a systematic bias component, and random error:

$$\text{logit}(p_{it}) = \text{logit}(\tilde{p}_{it}) - b_{it} + \zeta_{it}, \quad (9)$$

where b_{it} denotes the systematic bias.¹⁶ ζ_{it} denotes idiosyncratic *i.i.d.* errors with mean 0

¹⁶We use $-b_{it}$ instead of $+b_{it}$ so that the sign of the bias term is consistent with our logistic regressions in Section 5.3.

and variance ω^2 . Further, we parameterize the systematic bias b_{it} as follows:

$$b_{it} = (\alpha + s_i\beta)e^{-\delta(t-t_i^0)-\gamma V_{it}}. \quad (10)$$

Note that the parametrization of the bias is comprised of two terms: an additive term $(\alpha + s_i\beta)$, which, as before, specifies the direction and magnitude of the bias immediately after the first goal as a function of surprise (underreaction when > 0 ; overreaction when < 0), and a multiplicative term $e^{-\delta(t-t_i^0)-\gamma V_{it}}$, which estimates the extent to which the systematic bias in H1 is moderated by time lag ($e^{-\delta(t-t_i^0)}$) and by higher transaction volume ($e^{-\gamma V_{it}}$). A positive value for δ and γ indicates that the bias is attenuated by longer time lag (H2) and higher transaction volume (H3), respectively.

Finally, we specify a set of weakly informative priors for our model parameters. Specifically, we specify a $U(0, 1)$ prior distribution on $\tilde{p}_{it_i^r}$, an independent and diffuse $N(0, 100^2)$ distribution for the focal parameters $\alpha, \beta, \delta, \gamma$, and weakly informative $Inv-\chi^2(0.001, 1)$ distribution for the variance parameters $\sigma^2, \tau^2, \omega^2$ (Gelman et al., 2003). We then proceed to sample from the posterior distribution of all model parameters using Markov Chain Monte Carlo (MCMC) method (Johannes and Polson, 2009); the details of our MCMC procedure, which uses the forward filtering, backward sampling (FFBS) algorithm (Carter and Kohn, 1994), are included in Technical Appendix II. We run our MCMC sampler for 2,000 iterations, and discard the first 1,000 as a burn-in sample. The posterior mean and 90%, 95%, and 99% posterior intervals are summarized in Table 5.

The results in Table 5 confirm most of our earlier findings. First, the posterior means of α and β are 0.231 and -0.413 , respectively, which implies that bettors in general underreact to the first goal, but when the first goal is very surprising ($s_i > 0.56$), overreaction would occur instead. Note that the results for H1 are generally consistent with those shown in Section 5.2 and 5.3 (underreaction to most goals; overreaction when s_i is large), suggesting that the test of H1 is quite robust to the choice of statistical test and the level of data (match-level,

minute-by-minute, second-by-second) we look at.

Further, the time-decay parameter δ is estimated to be around 0.009, indicating that, consistently with H2, the systematic bias in H1 decreases over time ($\delta > 0$). For every minute after the first goal, the bias is reduced by about 40%. By the end of the fifth minute, the bias should decay to around 10% of its original magnitude; this is roughly consistent with the findings in Section 5.3 that the bias more or less vanishes six minutes after the first goal, and suggests the potential existence of arbitrageurs that provides a correction mechanism. However, we find that the posterior mean of γ is very close to 0, with a 90% posterior interval that covers zero. This suggests that H3 is not supported and that the bias in H1 is largely independent of transaction volume, perhaps because rational traders break up their orders and hide their activity.¹⁷

5.5 Economic Evidence

Having established the statistical evidence in support of H1 and H2, we now explore the economic significance of the over- and underreactions by developing a betting strategy that exploits such biases. First, we define the following simple strategies at two minutes after the first goal, depending on the level of surprise of the first goal:

- Back the outcome “scoring team wins” if $s_i < -c$
- Lay the outcome “scoring team wins” if $s_i > c$

Similar in spirit to Section 5.2, we vary c from 0.0 to 0.6 and calculate the profits for different values of c . The idea is to bet on (against) the scoring team when a goal is expected (surprising), and we try to identify at what value of c the strategies would start to make

¹⁷As stated in Section 4, our data are aggregated at the second-by-second level. Although we cannot find evidence to support H3, it is possible that the relationship between volume and the bias is detectable if transactions data are available.

profits. In each match, we back or lay an amount so that the payoff is £1 or -£1 if the scoring team wins, i.e., we bet an amount of £1/odds(scoring team wins).

We explore the profitability of each trading strategy by looking at the profit/loss that is generated (Total Earnings), as well as the profits divided by the volume (Earnings (%)). The results are shown in Table 6. The next two columns (Total Net Earnings and Net Earnings (%)) take into account the commissions Betfair charges: 5% of net winnings in each match (0% if loses money). The p -values shown in the last column are computed using non-parametric bootstrap, as the probability of each match is different and the distribution is not normal.

As can be seen in Table 6, all back strategies for $s_i < -c$ are profitable, with earnings that range from 1.98% to 4.88%, and net earnings from 1.33% to 4.42% (most of these are statistically significant). This implies that the odds of the scoring team show underreaction in these cases, consistent with our earlier evidence. For lay strategies (when $s_i > c$), when c is between 0.0 and 0.3, the net earnings of the strategies are negative; however, a profitable trading strategy using surprising goals could be developed by having $c > 0.4$. The net earnings after commissions of these profitable lay strategies are between 6.11% and 20.80%. Again, this is consistent with the results in Sections 5.2–5.4, where overreaction is observed when the goals are very surprising (when s_i is above a threshold). Note, however, that the net earnings of all lay strategies in Table 6 are statistically insignificant, perhaps because of the small number of matches in each strategy.

Table 7 constructs another set of strategies to exploit the biased reactions further, by backing (laying) when there is over-(under)reaction and using all matches in our sample. We also compare the strategies at different times after the goal. Specifically, we back the outcome “scoring team wins” when $s_i < 0.4$ (which includes both expected goals and moderately surprising goals) and lay the outcome “scoring team wins” when $s_i > 0.4$ (which includes the most surprising goals; this threshold is suggested by the results in Tables 3 and 6),

executed at 2, 3, and 6 minutes after the first goal.¹⁸ Not surprisingly, all these back and lay strategies are profitable after commissions, and the net profits of the combined strategies (back and lay) decrease over time: the net earnings go from 2.79% (p -value = 0.02) at 2 minutes, 1.85% (p -value = 0.07) at 3 minutes, to an insignificant 0.82% (p -value = 0.26) at 6 minutes.¹⁹ The profits seem high given the zero-sum game nature of the bets and the large volume in the window [+2, +6 minutes] (on average £152,000 per match) we observe in Table 1.²⁰

While our data do not provide bettors' identification, the profits could be seen as an estimate of the aggregate economic cost of the over- and underreactions we find. Considering the high volume and the fact that most match information is available instantly from the Internet, the strategies in Table 7 are economically significant. From the data, at the [+2, +3 minutes] window the total volume betting in the direction of our back and lay strategies exceeds £80 million; using the return at the three-minute mark (1.85%), the total profit is £80 million \times 1.85% = £1.48 million. This could represent a wealth transfer from irrational bettors to potential arbitrageurs who trade against the bias.

5.6 Comparison to Alternative Explanations

In this section, we explore several other potential interpretations of our results. We consider biases that may affect bettors' aggregate beliefs and behaviors, and show that our findings are inconsistent with these alternative explanations.

¹⁸Using $s_i < 0.56$ and $s_i > 0.56$ (suggested by the estimates in Table 5) gives similar results. These results are unreported.

¹⁹Again the net profits of the lay strategies are statistically insignificant. Nevertheless, by comparing the back only strategies and the combined strategies, we can see that the lay strategies help improve the net earnings, both in terms of the magnitude and p -value.

²⁰As a comparison, Gray and Gray (1997) report a strategy that earns over 4% after commissions in the National Football League (NFL) betting market. They note that the profits have gone down in the more recent part of their sample period.

5.6.1 Favorite–Longshot Bias

While we use the betting market to study reactions, we argue that our results using in-play odds are unlikely to be driven by the favorite–longshot bias. As discussed in Section 2.2, prior research in some betting markets documents this bias, where bettors tend to undervalue favorites and overvalue longshots.

Note that pre-match odds (odds before the start of the game) do not show similar results as the previous tables.²¹ Panel A of Table 8 shows that the Hosmer-Lemeshow test gives a test statistics of $\chi^2 = 23.2; p = 0.18$, which provides some evidence that pre-match odds are fairly efficient.

Panel B shows a logistic regression of match outcomes on pre-match odds-implied probability and the probability difference ($Implied-Pr_i(\text{team 1 wins}) - Implied-Pr_i(\text{team 2 wins})$), similar to equation (4) in Section 5.3.²² The intercept is much smaller compared to the logistic regression at 2 minutes after the goal (0.083 here vs. 0.157 in Table 3) and only marginally significant ($p\text{-value}=0.100$), while the coefficient of the probability difference is insignificant ($p\text{-value} = 0.655$). Therefore, we could not establish evidence that over- and underreactions occur before the match starts; suggesting the previous results using in-play odds are attributable to reactions to goals (i.e., adjustments in beliefs) rather than the favorite–longshot bias (prior beliefs).²³

²¹The sample size is bigger in this table because we do not filter matches based on the first goal.

²²We cannot calculate a surprise metric using the pre-match odds because there is not a goal yet. Instead, we construct the difference in probability measure using team 1 and team 2. For matches played at a team’s home ground, team 1 is the home team and team 2 is the away team. For other matches, teams 1 and 2 are randomly assigned.

²³One might consider studying the in-play odds *before* a goal is scored. However, this creates a bias because the analysis will be conditioning on a future event. For example, if we study the in-play odds 5 minutes before the first goal of the match, the sample will exclude matches that end with 0-0 but 0-0 is still a possibility when bettors place their bets.

5.6.2 Home Teams

Gray and Gray (1991) form a profitable strategy by betting on home teams in the National Football League (NFL) betting market. It is therefore also possible that any bias towards or against home teams could affect our results. In an unreported test, we add a dummy variable indicating the scoring team is the home team in our logistic regressions (equation (4)) in Table 3. This dummy variable is, however, not statistically significant in any of the regressions (from 2 minutes to 15 minutes); adding it also makes a poorer fit of the models based on the Akaike information criterion (AIC). Therefore, our results are unlikely to be affected by biases in the home team odds.

5.6.3 Unattractive Odds

One concern about the underreaction we document for expected goals is that the odds of the favorite team become even lower after the goal. This may suggest that the bets are no longer attractive to bettors, resulting in slow updating of the odds. Note that, however, Table 6 shows that even when the surprise metric s_i is smaller than -0.6 , betting on the favorite team at 2 minutes after the goal can still make a considerable amount of profits. We show further evidence here to argue that our previous findings are not driven by the low potential payoffs.

Table 9 repeats the logistic regressions (equation (4)) in Table 3, but dropping matches where the odds-implied probability of the scoring team is larger than 0.9 after the goal. The results are similar, and we conclude that the positive intercept in the regression is due to investors' underreaction.

6 Discussion and Conclusion

Using in-play soccer betting data, we show that over- and underreactions to information shocks depend on the level of surprise. The evidence is consistent with two behavioral biases: anchoring and salience. If the shock is expected, bettors tend to adjust too little from previous beliefs. A surprising shock, on the other hand, is likely to be salient and induces bettors' overweighting of its importance; as a result, underreaction decreases with surprise and overreaction occurs when the shock is very surprising. The biased reactions we find are corrected within five to six minutes after the shock. We also document economic significance by constructing a profitable strategy based on the reactions.

Our results shed light on how investors react to news. Fama (1998) argues that most asset pricing anomalies are “shaky”: they split randomly between over- and underreactions, and tend to disappear when alternative approaches are used to measure them. In contrast, consistent with behavioral predictions, we find evidence that over- and underreactions to new information are systematic and depend on the level of surprise. As Barberis, Shleifer, and Vishny (1998) argue, such evidence is difficult to be established in financial markets and “a real test... must await a better and more objective way of estimating the strength of news announcements.”²⁴

Based on our findings, we now outline three implications for asset pricing. First, we predict situations where overreaction should occur. Note that prior research typically associates news announcements with underreaction in financial markets, as there is usually post-event return continuation, i.e., average subsequent abnormal return has the same sign as the event-date stock price reaction (see Thaler, 1993, 2005; Hirshleifer, 2001; and papers

²⁴Our paper quantifies the degree of surprise of information shocks, and we use it as a proxy for the strength of the shocks. Papers in the finance literature often look at other aspects of information shocks, such as good news versus bad news, or one news announcement versus a series of news. It is, however, difficult to distinguish shocks in soccer games in similar ways: a goal scored is good news for the scoring team but bad news for the other team; there are not many goals in a game and each game only lasts around 90 minutes.

cited therein).²⁵ While our results are broadly consistent with these findings (we also find that there is underreaction to goals in general), we believe that there could be overreaction when the news is salient and surprising. However, as we argue earlier, an immediate question is how we quantify salience or surprise objectively. The degree of surprise of an event in financial markets could be measured differently; for example, Chan, Jegadeesh, and Lakonishok (1996) use three different proxies for earnings surprises: standardized unexpected earnings, cumulative abnormal stock return around the earnings announcement date, and changes in analysts' forecasts, but the correlations between the proxies are quite low, from 0.115 to 0.440. Therefore, like Barberis, Shleifer, and Vishny (1998), we argue that it is important to have an a priori way to measure the strength of a one-time shock. If such a measure captures investors' tendency to overweight salience, then we expect to see overreaction to some surprising events, as psychology theories and our paper suggest. This result will help us better understand investors' biased reactions to different events, as well as improve the profitability of strategies based on post-event return continuation (e.g., earnings momentum in Chan, Jegadeesh, and Lakonishok, 1996).

Second, stock price reactions are often used as predictors of binary outcomes, similar to our study. An example is mergers and acquisitions: Samuelson and Rosenthal (1986) and Luo (2005) argue that stock prices after the initial announcements are predictors of the ultimate success or failure of the deal. We think, however, there may be over- and underreactions at least immediately after the announcements, depending on how salient the news is. One could see whether investors' biased reactions affect the predictability of stock prices, or in the case of mergers and acquisitions, whether the biases could generate a feedback effect that changes the deal success probability: for example, if biases cause misvaluations of the merging firms that affect acquisitions in a way proposed by Shleifer and Vishny (2003).

²⁵As summarized by Hirshleifer (2001), events that show post-event return continuation include stock splits, tender offer and open market repurchases, equity carveouts, spinoffs, accounting write-offs, analyst earnings forecast revisions, analyst stock recommendations, dividend initiations and omissions, seasoned issues of debt and common stock, public announcement of previous insider trades, and venture capital share distributions.

Finally, our results may help us study the underlying mechanism of momentum. Behavioral theory models are yet to reach a consensus on whether momentum is caused by over- or underreactions to news. Da, Gurun, and Warachka (2011) show that momentum profits are lower when information arrives in large amounts at discrete timepoints than when information arrives continuously in small amounts. They claim that the former case attracts investor attention and decreases underreaction. If we interpret discrete information as more surprising or salient, then the implications of our results provide additional support that momentum is caused by underreaction; in cases where salient news is likely to decrease the extent of underreaction (or even trigger overreaction), price momentum is lower.

Our analysis and future research can therefore contribute by classifying situations where over- or underreactions is likely to occur using surprise and salience, as well as understanding stock prices' predictability of outcomes and anomalies by linking them to investor reactions. By combining in-play soccer bets with first goals, we create a dataset that has several advantages over traditional datasets, as we argue in the Introduction. The betting market allows researchers to examine apparent shocks and actual outcomes in a unique setting, and our data can be reused to study other investor behaviors and their impact on prices.

Technical Appendix

I. Computation of the bivariate Poisson process and the “efficient drift,” g_{it}

We briefly discuss how g_{it} , the “efficient” drift of odds under an assumed model of soccer goal scoring, is estimated. First, we begin by assuming that goal scoring by the two teams in a soccer match (denoted as x_{i1} and x_{i2} , respectively) follows an independent bivariate Poisson process (Maher, 1982) with rates λ_{i1} and λ_{i2} , respectively. Then, given the Poisson rate parameters $(\lambda_{i1}, \lambda_{i2})$, the current goal difference (z_{it}) and the running time in the match (t), the conditional probability that the scoring team will win at the end of the match can be computed numerically. We denote conditional probability (at time t) as $f(\lambda_{i1}, \lambda_{i2}, z_{it}, t)$. By definition, we compute the “efficient” drift in (logit-transformed) probability by:

$$g_{it} = \text{logit}(f(\lambda_{i1}, \lambda_{i2}, z_{it}, t)) - \text{logit}(f(\lambda_{i1}, \lambda_{i2}, z_{i,t-1}, t - 1)). \quad (\text{A1})$$

For concreteness, we provide a numerical example showing how g_{it} is computed. Let $\lambda_{i1} = 1.0$, $\lambda_{i2} = 0.7$, (we will discuss how $(\lambda_{i1}, \lambda_{i2})$ are estimated later), and that team 1 (the scoring team) is leading by 1-0 at $t = 1800$ (30 minutes into the match); assume that the total duration of a match, including injury time of both halves, is around 95 minutes (5700 seconds; excluding the halftime break). Because of the independent bivariate Poisson assumption, for the rest of the match, the number of goals that are scored by team 1 follows a Poisson distribution with rate $\frac{5700-1800}{5700}\lambda_{i1} = 0.68$, and the number of goals scored by team 2 follows a Poisson distribution with rate $\frac{5700-1800}{5700}\lambda_{i2} = 0.48$. From the probability mass function of the Poisson distribution, we can easily compute the conditional probability that team 1 will continue to win the match is $f(1.0, 0.7, 1, 1800) = 0.777496$.²⁶ Similarly, if no additional goal occurs in the next second, we can compute the conditional probability that team 1 win

²⁶This computation is coded up as a function in R. The R code is available from the authors upon request.

as $f(1.0, 0.7, 1, 1801) = 0.777517$. Clearly, because team 1 is in the lead, the passage of time without a goal increases the conditional probability that team 1 will win, resulting in a positive drift. To compute g_{it} , we have $g_{it} = \text{logit}(0.777517) - \text{logit}(0.777496) = +0.00012$.²⁷

The only computation left is the estimation of $(\lambda_{i1}, \lambda_{i2})$. We calibrate the $(\lambda_{i1}, \lambda_{i2})$ parameters using the pre-match implied probabilities. That is, the value of $(\lambda_{i1}, \lambda_{i2})$ is chosen such that the (team 1 win, draw, team 2 win) probabilities are in maximal agreement with the pre-match implied probabilities. The details, including the R code used for this calibration, are available from the authors upon request.

Finally, we note that one may argue that the assumption of an independent bivariate Poisson scoring process is an over-simplification of an actual soccer match. Indeed, many refinements and elaborations that build upon on the bivariate Poisson model have been proposed in the literature (e.g., Dixon and Coles, 1997; Karlis and Ntzoufras, 2003); most of these refinements could potentially allow the function f to capture the “true” conditional probability more accurately. For our purpose, we prefer to use the parsimonious independent Poisson assumption for two reasons. First, since we use only the first difference of f (in logit scale) as an additional term in the prior distribution of π_{it} , the difference between the models is likely to be minimal. Second, in equation (7) of our model we have already included the idiosyncratic error term ε_{it} to capture any modeling error; our estimation shows that the size of the error term ε_{it} tends to be very small (see Table 5), indicating that the bivariate Poisson distribution provides an adequate description of the evolution of the “true” conditional probability. Finally, the independent bivariate Poisson assumption provides vast computational advantage over more elaborate models; specifically, it allows us to calibrate $(\lambda_{i1}, \lambda_{i2})$ in a straightforward manner. In contrast, it is not clear how the parameters of

²⁷If a goal is scored after 40 minutes in the first half, the [+2, +5 min] window we analyze may cover the halftime break, where g_{it} should be very close to zero as there is little amount of information. However, from the Betfair and Soccerbet data we are unable to pinpoint the time when the halftime break starts. We do not analyze these goals differently and believe that the impact should be minimal: given that our estimate of g_{it} is typically very small and that only 137 (6.8%) of the 2,017 goals in our sample are scored after 40 minutes in the first half.

more elaborate models can be estimated using only the pre-match odds in our data.

II. Markov Chain Monte Carlo (MCMC) Sampling

We outline a standard Markov Chain Monte Carlo (MCMC) sampling procedure that we use to sample from the posterior distribution of the model parameters (Johannes and Polson, 2009). The details of our computation procedure, including the R code used to estimate the model, are available from the authors upon request. In each MCMC iteration, we draw from the full conditional distribution of each model parameter using the following order:

1. *Drawing from the full conditional distribution of \tilde{p}_{it^r}* : A Gaussian random-walk Metropolis-Hastings algorithm (Hastings 1970) is used to draw $\text{logit}(\tilde{p}_{it^r})$. The scale of the Gaussian random walk proposal distribution is tuned to achieve an acceptance rate of around 40% (Gelman et al. 2003).
2. *Drawing from the full conditional distribution of \tilde{p}_{it} ($t > t_i^r$)*: For $\text{logit}(\tilde{p}_{it})$ other than the first period, the Gaussian form of equations (7)–(9) allows us to sample from its full conditional distribution using the forward filtering, backward sampling (FFBS) algorithm (Carter and Kohn, 1994). Essentially, we compute the conditional mean and variances of $\text{logit}(\tilde{p}_{it})$ for each t using a modified Kalman filter, then sample all $\text{logit}(\tilde{p}_{it})$'s at the same time. This allows us to sample from the full conditional distribution of $\text{logit}(\tilde{p}_{it})$ very efficiently (Carter and Kohn, 1994).
3. *Drawing from the full conditional distribution of $(\alpha, \beta, \delta, \gamma)$* : We again use a random walk Metropolis-Hastings algorithm to sample from the full conditional distribution of $(\alpha, \beta, \delta, \gamma)$. We sample each parameter one-by-one, and adapt the scale of the random-walk proposal distribution to achieve an acceptance rate of around 40% (Gelman et al., 2003).

4. *Drawing from the full conditional distribution of $(\sigma^2, \tau^2, \omega^2)$* : Finally, because the variance parameters are given conjugate, weakly informative $Inv\text{-}\chi^2(0.001, 1)$ priors, we sample from the full conditional distribution of $(\sigma^2, \tau^2, \omega^2)$ using standard conjugate computations (see Gelman et al., 2003).

We repeat the above procedure for 2,000 iterations, and discarded the first 1,000 draws as a burn-in sample. The last 1,000 draws are used to summarize the posterior distribution of model parameters by computing their posterior means and posterior intervals.

References

- Arrow, Kenneth J., Robert Forsythe, Michael Gorham, Robert Hahn, Robin Hanson, John O. Ledyard, Saul Levmore, Robert Litan, Paul Milgrom, Forrest D. Nelson, George R. Neumann, Marco Ottaviani, Thomas C. Schelling, Robert J. Shiller, Vernon L. Smith, Erik Snowberg, Cass R. Sunstein, Paul C. Tetlock, Philip E. Tetlock, Hal R. Varian, Justin Wolfers, and Eric Zitzewitz, 2008, The promise of prediction markets, *Science* 320, 877–878.
- Asch, Peter, Burton G. Malkiel, and Richard E. Quandt, 1984, Market efficiency in race-track betting, *Journal of Business* 57, 165–175.
- Avery, Christopher and Judith Chevalier, 1999, Identifying investor sentiment from price paths: The case of football betting, *Journal of Business* 72, 493–521.
- Barberis, Nicholas, Andrei Shleifer, and Robert Vishny, 1998, A model of investor sentiment, *Journal of Financial Economics* 49, 307–343.
- Barclay, Michael J. and Jerold B. Warner, 1993, Stealth trading and volatility: Which trades move prices? *Journal of Financial Economics* 34, 281–305.
- Bloomfield, Robert, Robert Libby, and Mark Nelson, 2000, Underreactions, overreactions and moderated confidence, *Journal of Financial Markets* 3, 113–137.
- Carter, C.K., and R. Kohn, 1994, On Gibbs Sampling for state space models, *Biometrika* 81, 541–553.
- Chan, Louis K.C., Narasimhan Jegadeesh, and Josef Lakonishok, 1996, Momentum strategies, *Journal of Finance* 51, 1681–1713.
- Croxson, Karen and J. James Reade, 2011, Information and efficiency: Goal arrival in soccer betting, Working paper, University of Birmingham.
- Da, Zhi, Umit G. Gurun, and Mitch Warachka, 2011, Frog in the pan: Continuous information and momentum, Working paper, University of Notre Dame, University of Texas at Dallas, and Singapore Management University.
- Daniel, Kent, David Hirshleifer, and Avanidhar Subrahmanyam, 1998, Investor psychology and security market under- and overreactions, *Journal of Finance* 53, 1839–1885.
- De Long, J. Bradford, Andrei Shleifer, Lawrence Summers, and Robert J. Waldmann, 1990, Positive feedback investment strategies and destabilizing rational speculation, *Journal of Finance* 45, 375–395.
- Dixon, Mark J. and Stuart G. Coles, 1997, Modelling association football scores and inefficiencies in the football betting market, *Journal of Royal Statistical Association (Series C)* 46, 265–280.

- Durham, Gregory R., Michael G. Hertzel, and J. Spencer Martin, 2005, The market impact of trends and sequences in performance: New evidence, *Journal of Finance* 60, 2551–2569.
- Durham, Gregory R. and Mukunthan Santhanakrishnan, 2008, The relative importance of strength and weight in processing new information in the college football betting market, *Journal of Prediction Markets* 2, 13–28.
- Easley, David and Maureen OHara, 1987, Price, trade size, and information in securities markets, *Journal of Financial Economics* 19, 69–90.
- Engle, Robert F., 2000, The Econometrics of ultra-high-frequency data, *Econometrica* 68, 1–22.
- Fama, Eugene F., 1998, Market efficiency, long-term returns, and behavioral finance, *Journal of Financial Economics* 49, 283–306.
- Frazzini, Andrea, 2006, The disposition effect and underreaction to news, *Journal of Finance* 61, 2017–2046.
- Gandar, John, Richard Zuber, Thomas OBrien, and Ben Russo, 1988, Testing rationality in the point spread betting market, *Journal of Finance* 43, 995–1008.
- Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin, 2003, *Bayesian Data Analysis, 2nd Edition*, Chapman & Hall.
- Gil, Ricard and Steven D. Levitt, 2007, Testing the efficiency of markets in the 2002 World Cup, *Journal of Prediction Markets* 1, 255–270.
- Gneezy, Uri, 1996, Probability judgment in multi-stage problems: Experimental evidence of systematic biases, *Acta Psychologica* 93, 59–68.
- Golec, Joseph and Maurry Tamarkin, 1991, The degree of inefficiency in the football betting market, *Journal of Financial Economics* 30, 311–323.
- Gray, Philip K. and Stephen F. Gray, 1997, Testing market efficiency: Evidence from the NFL sports betting market, *Journal of Finance* 52, 1725–1737.
- Griffin, Dale and Amos Tversky, 1992, The weighing of evidence and the determinants of confidence, *Cognitive Psychology* 24, 411–435.
- Grinblatt, Mark, and Bing Han, 2005, Prospect theory, mental accounting, and momentum, *Journal of Financial Economics* 78, 311–339.
- Hasbrouck, Joel, 1988, Trades, quotes, inventories, and information, *Journal of Financial Economics* 22, 229–252.
- Hasbrouck, Joel, 1991, Measuring the information content of stock trades, *Journal of Finance* 46, 179–207.

- Hastings, W.K., 1970, Monte Carlo sampling methods using Markov Chains and their applications, *Biometrika* 57, 97–109.
- Hirshleifer, David, 2001, Investor psychology and asset pricing, *Journal of Finance* 56, 1533–1597.
- Hong, Harrison and Jeremy C. Stein, 1999, A unified theory of underreaction, momentum trading, and overreaction in asset markets, *Journal of Finance* 54, 2143–2184.
- Hosmer, David W. and Stanley Lemeshow, 2000, *Applied Logistic Regression*, New York: Wiley.
- Itti, Laurent and Pierre Baldi, 2009, Bayesian surprise attracts human attention, *Vision Research* 49, 1295–1306.
- Jegadeesh, Narasimhan and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *Journal of Finance* 48, 65–91.
- Johannes, Michael and Nick Polson, 2009, MCMC Methods for Financial Econometrics, in *Handbook of Financial Econometrics*, Yacine Ait-Sahalia and Lars Peter Hansen (Eds), 2, 1–72, North Holland.
- Kahneman, Daniel, Paul Slovic, and Amos Tversky, 1982, *Judgment Under Uncertainty: Heuristics and Biases*, Cambridge University Press.
- Karlis, Dimitris and Ioannis Ntzoufras, 2003, Analysis of sports data using bivariate Poisson models, *Journal of the Royal Statistical Association (Series D)* 52, 381–393.
- Klibanoff, Peter, Owen Lamont, and Thierry A. Wizman, 1998, Investor reaction to salient news in closed-end country funds, *Journal of Finance* 53, 673–699.
- Levitt, Steven D. and John A. List, 2007a, What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives* 21, 153–174.
- Levitt, Steven D. and John A. List, 2007b, Viewpoint: On the generalizability of lab behaviour to the field, *Canadian Journal of Economics* 40, 347–370.
- Luo, Yuanzhi, 2005, Do insiders learn from outsiders? Evidence from mergers and acquisitions, *Journal of Finance* 60, 1951–1982.
- Maher, M. J., 1982, Modelling association football scores, *Statistica Neerlandica* 36, 109–118.
- Michaely, Roni, Richard H. Thaler, and Kent L. Womack, 1995, Price reactions to dividend initiations and omissions: Overreaction or drift? *Journal of Finance* 50, 573–608.
- Nelder, J. A. and R.W.M. Wedderburn, 1972, Generalized linear models, *Journal of Royal Statistical Society (Series A)* 135, 370–384.

- Poteshman, Allen M., 2001, Underreaction, overreaction, and increasing misreaction to information in the options market, *Journal of Finance* 56, 851–876.
- Ramalingegowda, Santhosh, Tao Shu, and P. Eric Yeung, 2011, Moderated confidence and under- and overreactions in the stock markets, Working paper, University of Georgia.
- Ranganath, Charan and Gregor Rainer, 2003, Neural mechanisms for detecting and remembering novel events, *Nature Reviews Neuroscience* 4, 193–202.
- Samuelson, William and Leonard Rosenthal, 1986, Price movements as indicators of tender offer success, *Journal of Finance* 41, 481–499.
- Shleifer, Andrei and Robert W. Vishny, 2003, Stock market driven acquisitions, *Journal of Financial Economics* 70, 295–311.
- Thaler, Richard H., 1993, *Advances in Behavioral Finance*, Russell Sage Foundation.
- Thaler, Richard H., 2005, *Advances in Behavioral Finance, Volume II (Roundtable Series in Behavioral Economics)*, Princeton University Press.
- Thaler, Richard H. and William T. Ziemba, 1988, Anomalies: Parimutuel betting markets: Racetracks and lotteries, *Journal of Economic Perspectives* 2, 167–174.
- Tversky, Amos and Daniel Kahneman, 1974, Judgments under uncertainty: Heuristics and biases, *Science* 185, 1124–1131.
- Vergin, Roger C., 2001, Overreaction in the NFL point spread market, *Applied Financial Economics* 11, 497–509.
- Wolfers, Justin and Eric Zitzewitz, 2004, Prediction markets, *Journal of Economic Perspectives* 18, 107–126.
- Wolfers, Justin and Eric Zitzewitz, 2006, Interpreting prediction market prices as probabilities, Working paper, University of Pennsylvania and Dartmouth College.
- Wright, William F. and Urton Anderson, 1989, Effects of situation familiarity and financial incentives on use of the anchoring and adjustment heuristic for probability assessment, *Organizational Behavior and Human Decision Processes* 44, 68–82.

Table 1
Match Statistics

This table shows the match statistics for our sample of 2,017 matches. *Matched Volume* is the volume (in British pounds) that are matched by Betfair. Betfair matches back and lay orders submitted by bettors. The volume is calculated for all in-play bets and bets placed between 2 minutes and 6 minutes after the first goal. *First Goal* is the time (in minutes) of the first goal after the start of the match. *Scoring Team Is Favorite* and *Scoring Team Wins* are dummy variables, indicating that the first scoring team is the favorite team (identified using odds one minute before the first goal) and wins the match, respectively. *Odds-Implied Probability* is the odds-implied probability of the scoring team, measured at different 2, 3, and 6 minutes after the first goal. The odds-implied probability is the reciprocal of the odds, divided by the sum of the reciprocal of the odds of all outcomes.

	Mean	Median	Standard Deviation	Minimum	Maximum
Matched Volume (In-play, £)	1,481,627	917,054	1,627,070	9,654	13,400,407
Matched Volume ([+2,+6 minutes] after First Goal, £)	152,523	93,779	175,621	103	1,708,552
First Goal (Minutes)	35.84	25.97	30.47	0.18	113.07
Scoring Team Is Favorite	0.670	1	0.470	0	1
Scoring Team Wins	0.712	1	0.453	0	1
Odds-Implied Probability (+2 minutes after First Goal)	0.686	0.720	0.187	0.017	0.994
Odds-Implied Probability (+3 minutes after First Goal)	0.691	0.726	0.190	0.002	0.995
Odds-Implied Probability (+6 minutes after First Goal)	0.694	0.730	0.192	0.034	0.997

Table 2
Descriptive Analysis: Actual versus Expected Number of Wins

This table shows the actual and expected number of wins for different values of c . The left-hand panel shows matches with surprise metric $s_i < -c$, while the right-hand panel shows matches with surprise $s_i > c$. The metric s_i is defined as the odds-implied probability of the non-scoring team minus that of the scoring team, measured at one minute before the first goal of the match. The odds-implied probability is the reciprocal of the odds, divided by the sum of the reciprocal of the odds of all outcomes. The expected number of wins is estimated using the null hypothesis that the odds-implied probability is the true probability (see main text for a full description). The last column contains the p-value of the test statistic (generated by simulations) for the difference between the expected and actual number of wins. *, **, and *** denote differences that are 10%, 5%, and 1% significant, respectively.

c	Favorite scores: surprise $< -c$				Underdog scores: surprise $> c$			
	N	Expected # of Wins	Actual # of Wins	p-value	N	Expected # of Wins	Actual # of Wins	p-value
0	1299	1014.27	1067	0.000***	638	313.84	320	0.312
0.1	1090	868.50	918	0.000***	477	213.97	212	0.451
0.2	860	704.51	735	0.003***	318	123.35	124	0.505
0.3	664	558.26	578	0.016**	224	75.88	78	0.392
0.4	484	416.58	428	0.077*	145	41.73	38	0.288
0.5	336	295.84	306	0.040**	95	23.81	21	0.302
0.6	191	173.44	183	0.006***	56	11.86	9	0.219

Table 3
Estimation of Bias After First Goal:
Logistic Regressions of Match Outcomes

This table shows the results of the logistic regressions of match outcomes. The dependent variable is 1 if the scoring team wins and 0 otherwise. The independent variables are the odds-implied probability of the scoring team, measured at different times after the first goal, as well as the surprise metric, s_t . The coefficient of the odds-implied probability is set to be 1, and the coefficient of the metric s_t is β . α represents the intercept. The metric s_t is defined as the odds-implied probability of the non-scoring team minus that of the scoring team, measured at one minute before the first goal of the match. The odds-implied probability is the reciprocal of the odds, divided by the sum of the reciprocal of the odds of all outcomes. *, **, and *** denote 10%, 5%, and 1% significance, respectively.

Time (Minutes After First Goal)	Intercept (α)	p-value ($\alpha = 0$)	Coeff. of s_t (β)	p-value ($\beta = 0$)	p-value ($\alpha = \beta = 0$)
2	0.157***	0.005	-0.385**	0.029	0.001***
3	0.109**	0.051	-0.291*	0.095	0.019**
4	0.086	0.122	-0.294*	0.092	0.042**
5	0.073	0.193	-0.285	0.104	0.073*
6	0.060	0.285	-0.279	0.112	0.112
7	0.054	0.337	-0.229	0.192	0.208
8	0.055	0.334	-0.236	0.180	0.196
9	0.043	0.453	-0.221	0.209	0.282
10	0.038	0.511	-0.221	0.214	0.313
11	0.034	0.553	-0.253	0.157	0.254
12	0.025	0.665	-0.274	0.126	0.238
13	0.020	0.733	-0.269	0.135	0.267
14	0.026	0.655	-0.263	0.150	0.270
15	0.013	0.829	-0.267	0.144	0.302

Table 4
Estimation of Bias After First Goal: Hosmer-Lemeshow Test

Hosmer-Lemeshow test statistic is shown in the table, for the in-play odds at different times after the first goal is scored. The Hosmer-Lemeshow test, which compares a probability assessment versus actual outcomes, is a test for the goodness-of-fit for logistic regression models. The test statistic follows a chi-square($k-2$) distribution, where k is the number of “bins” used to group the data (chosen to be $k = 20$ here) (see, for example, Hosmer and Lemeshow, 2000). * and ** denote 10% and 5% significance, respectively.

Time (Minutes After First Goal)	H-L test stat	p-value
2	32.35**	0.02
3	27.46*	0.07
4	26.45*	0.09
5	16.64	0.55
6	23.30	0.18
7	21.97	0.23
8	22.81	0.20
9	22.16	0.22
10	23.78	0.16
11	13.82	0.74
12	17.95	0.46
13	22.09	0.23
14	18.78	0.41
15	23.96	0.16

Table 5
Estimation of True Probability and Bias After First Goal:
Bayesian Model

We set up a formal Bayesian framework to model the in-play odds-implied probability as the sum of the true unobserved probability, a systematic bias component, and random error. The systematic bias depends on an intercept term and the surprise metric s_b , as well as betting volume and time elapsed after the goal. The error in modeling the true probability (equation (7)) has a variance of σ^2 in the absence of another major event (another goal, penalty awarded, or red card), and τ^2 if another major event has occurred. The random error has a variance of ω^2 . We use a Markov Chain Monte Carlo (MCMC) procedure to sample from the posterior distribution of all parameters. See Section 5.4 of the main text for the model and Technical Appendix for details. 90%, 95%, and 99% posterior intervals are shown. *** denotes coefficients that are 1% significant.

		Estimate	Posterior Intervals		
			90%	95%	99%
Intercept	α	0.231***	(0.226, 0.235)	(0.225, 0.235)	(0.225, 0.235)
Coeff. of s_i	β	-0.413***	(-0.422, -0.404)	(-0.423, -0.403)	(-0.425, -0.402)
Coeff. of time elapsed	δ	0.009***	(0.009, 0.009)	(0.009, 0.009)	(0.009, 0.009)
Coeff. of volume	γ	0.000	(-0.001, 0.001)	(-0.001, 0.001)	(-0.002, 0.001)
Model Error (No Event)	σ^2 (x10 ³)	0.880	(0.872, 0.891)	(0.870, 0.893)	(0.867, 0.895)
Model Error (Event Occurred)	τ^2	3.437	(2.745, 4.335)	(2.605, 4.505)	(2.434, 5.068)
Random Error	ω^2 (x10 ³)	0.199	(0.192, 0.205)	(0.192, 0.206)	(0.191, 0.208)

Table 6
Trading Strategies at Two Minutes After First Goal

Trading strategies are formed based on the surprise metric s_i . The top panel shows back strategies: bet on scoring team two minutes after the first goal if $s_i < -c$. The bottom panel shows lay strategies: bet against the scoring team two minutes after the first goal if $s_i > c$. For each match, the amount we back or lay is given by $1/\text{odds}$. The *Total Volume* (total dollar amount of bets across all matches), *Total Earnings* (total dollar amount of earnings across all matches), *Earnings (%)* (total earnings divided by total volume), *Total Net Earnings* (total earnings after commissions), and *Net Earnings (%)* (total earnings after commissions divided by total volume) are shown for all strategies. The last column shows the p-value for *Net Earnings (%)*, estimated by bootstrapping. *, **, *** denote *Net Earnings (%)* that are 10%, 5%, and 1% significant, respectively.

Back: surprise < -c							
c	N	Total Volume (£)	Total Earnings (£)	Earnings (%)	Total Net Earnings (£)	Net Earnings (%)	p-value (Net Earnings = 0)
0	1299	1022.88	44.123	4.31%	33.379	3.26%***	0.002
0.1	1090	875.29	42.713	4.88%	34.025	3.89%***	0.001
0.2	860	709.66	25.337	3.57%	19.160	2.70%**	0.029
0.3	664	562.21	15.785	2.81%	11.504	2.05%*	0.086
0.4	484	419.70	8.298	1.98%	5.584	1.33%	0.183
0.5	336	297.73	8.270	2.78%	6.603	2.22%*	0.092
0.6	191	174.50	8.504	4.87%	7.714	4.42%**	0.012
Lay: surprise > c							
c	N	Total Volume (£)	Total Earnings (£)	Earnings (%)	Total Net Earnings (£)	Net Earnings (%)	p-value (Net Earnings = 0)
0	638	316.57	-3.434	-1.08%	-10.640	-3.36%	0.173
0.1	477	215.28	3.280	1.52%	-2.301	-1.07%	0.395
0.2	318	124.32	0.320	0.26%	-3.262	-2.62%	0.330
0.3	224	76.37	-1.634	-2.14%	-3.999	-5.24%	0.291
0.4	145	42.10	4.098	9.73%	2.573	6.11%	0.304
0.5	95	24.03	3.031	12.61%	2.100	8.74%	0.279
0.6	56	12.01	3.013	25.08%	2.499	20.80%	0.171

Table 7
Trading Strategies Based on Over- and Underreactions

Trading strategies are formed based on the surprise metric s_i at different times after the first goal. The top panel shows back strategies: bet on scoring team after the first goal if $s_i < 0.4$. These are matches that tend to show underreaction, based on previous results. The middle panel shows lay strategies: bet against the scoring team after the first goal if $s_i > 0.4$. These are matches that tend to show overreaction, based on previous results. For each match, the amount we back or lay is given by $1/\text{odds}$. The bottom panel combines the two strategies. The *Total Volume* (total dollar amount of bets across all matches), *Total Net Earnings* (total earnings after commissions), and *Net Earnings (%)* (total earnings after commissions divided by total volume) are shown for all strategies. The last column shows the p-value for *Net Earnings (%)*, estimated by bootstrapping. * and ** denote *Net Earnings (%)* that are 10% and 5% significant, respectively.

Back: surprise < 0.4					
Time (Minutes After First Goal)	N	Total Volume (£)	Total Net Earnings (£)	Net Earnings (%)	p-value (Net Earnings = 0)
2	1792	1297.35	34.853	2.69%**	0.028
3	1833	1330.78	23.443	1.76%*	0.087
6	1827	1333.71	8.388	0.63%	0.328
Lay: surprise > 0.4					
Time (Minutes After First Goal)	N	Total Volume (£)	Total Net Earnings (£)	Net Earnings (%)	p-value (Net Earnings = 0)
2	145	42.10	2.573	6.11%	0.304
3	147	42.54	2.028	4.77%	0.346
6	148	43.42	2.888	6.65%	0.298
Combined: Back when surprise < 0.4 and Lay when surprise > 0.4					
Time (Minutes After First Goal)	N	Total Volume (£)	Total Net Earnings (£)	Net Earnings (%)	p-value (Net Earnings = 0)
2	1937	1339.45	37.426	2.79%**	0.018
3	1980	1373.32	25.472	1.85%*	0.073
6	1975	1377.13	11.276	0.82%	0.260

Table 8
Comparison: Analyses Using Pre-Match Odds

Hosmer-Lemeshow test statistic is shown in Panel A of the table for the pre-match odds. The Hosmer-Lemeshow test, which compares a probability assessment versus actual outcomes, is a test for the goodness-of-fit for logistic regression models. The test statistic follows a chi-square($k-2$) distribution, where k is the number of “bins” used to group the data (chosen to be $k = 20$ here) (see, for example, Hosmer and Lemeshow, 2000).

Panel B shows the results of the logistic regression of match outcomes. The dependent variable is 1 if the team 1 wins and 0 otherwise (team 1 = home team or random team). The independent variables are the odds-implied probability of team 1, measured at the start of the match, as well as the *Probability Difference*. The coefficient of the odds-implied probability is set to be 1, and α represents the intercept. The *Probability Difference* is defined as the odds-implied probability of team 1 minus that of team 2, measured at the start of the match. The odds-implied probability is the reciprocal of the odds, divided by the sum of the reciprocal of the odds of all outcomes.

Panel A: Hosmer-Lemeshow Test		
Pre-match	H-L test stat	p-value
	23.20	0.180

Panel B: Logistic Regression		
	Estimate	p-value
Intercept α	0.0830	0.100
Coefficient of Probability Difference	0.0601	0.655

Table 9
Comparison: Logistic Regressions of Match Outcomes
(Dropping Matches With Probability > 0.9)

This table shows the results of the logistic regressions of match outcomes, dropping matches where the odds-implied probability of the scoring team > 0.9. The dependent variable is 1 if the scoring team wins and 0 otherwise. The independent variables are the odds-implied probability of the scoring team, measured at different times after the first goal, as well as the surprise metric, s_t . The coefficient of the odds-implied probability is set to be 1, and the coefficient of the metric s_t is β . α represents the intercept. The metric s_t is defined as the odds-implied probability of the non-scoring team minus that of the scoring team, measured at one minute before the first goal of the match. The odds-implied probability is the reciprocal of the odds, divided by the sum of the reciprocal of the odds of all outcomes. *, **, and *** denote 10%, 5%, and 1% significance, respectively.

Time (Minutes After First Goal)	Intercept (α)	p-value ($\alpha = 0$)	Coeff. of s_t (β)	p-value ($\beta = 0$)	p-value ($\alpha = \beta = 0$)
2	0.153***	0.007	-0.406**	0.022	0.000***
3	0.104*	0.061	-0.313*	0.074	0.017**
4	0.082	0.142	-0.316*	0.071	0.038**
5	0.068	0.221	-0.307*	0.081	0.066*
6	0.056	0.321	-0.301*	0.088	0.099*
7	0.050	0.378	-0.251	0.156	0.190
8	0.050	0.375	-0.258	0.145	0.178
9	0.038	0.502	-0.243	0.171	0.256
10	0.033	0.564	-0.243	0.174	0.281
11	0.030	0.607	-0.276	0.125	0.223
12	0.020	0.724	-0.298*	0.099	0.204
13	0.015	0.794	-0.293	0.106	0.227
14	0.022	0.714	-0.286	0.118	0.233
15	0.008	0.893	-0.291	0.114	0.256

Figure 1
Histogram of Surprise Metric

This figure shows the distribution of the surprise metric, s_t . The metric s_t is defined as the odds-implied probability of the non-scoring team minus that of the scoring team, measured at one minute before the first goal of the match. The odds-implied probability is the reciprocal of the odds, divided by the sum of the reciprocal of the odds of all outcomes.

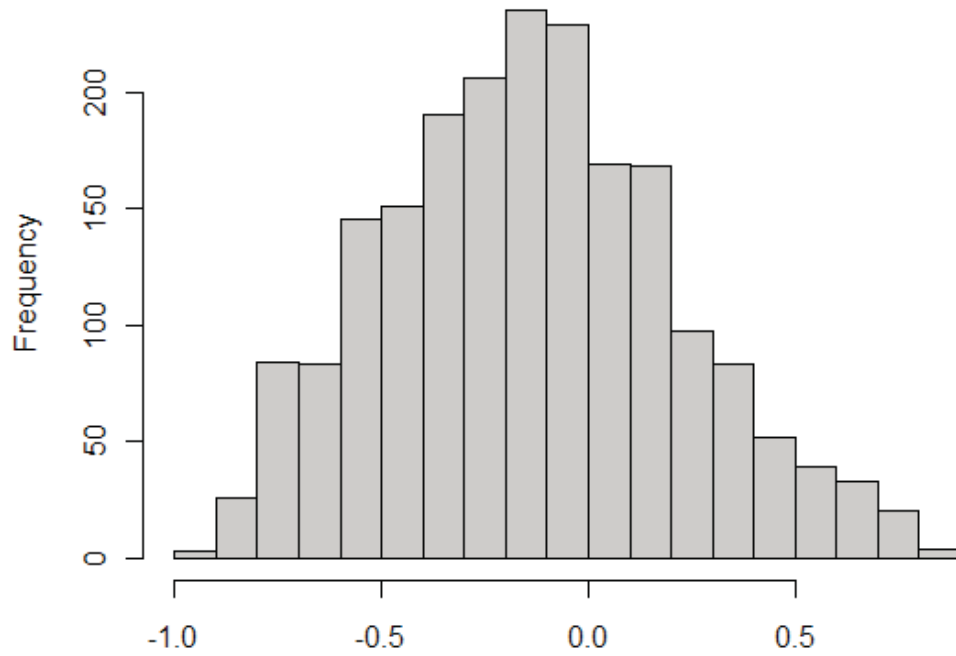


Figure 2
Reaction After First Goal: Two Examples

This figure shows the *Odds-Implied Probability* of the scoring team for two matches in our sample: Italy vs. New Zealand (New Zealand is the scoring team and the underdog before the goal) and Arsenal vs. Watford (Arsenal is the scoring team and the favorite before the goal). The graphs show the odds-implied probabilities around the first goal of the match. The odds-implied probability is the reciprocal of the odds, divided by the sum of the reciprocal of the odds of all outcomes.

