University of Zurich[UZH]

Department of Business Administration

UZH Business Working Paper Series

(ISSN 2296-0422)

Working Paper No. 357

The Importance of Suspense and Surprise in Entertainment Demand: Evidence from Wimbledon

Paolo Bizzozero, Raphael Flepp and Egon Franck

July 2016

University of Zurich, Plattenstrasse 14, CH-8032 Zurich,
http://www.business.uzh.ch/forschung/wps.html

UZH Business Working Paper Series

Contact Details

Paolo Bizzozero

University of Zurich

Department of Business Administration

paolo.bizzozero@business.uzh.ch

Raphael Flepp

University of Zurich

Department of Business Administration

raphael.flepp@business.uzh.ch

Egon Franck

University of Zurich

Department of Business Administration

egon.franck@business.uzh.ch

The importance of suspense and surprise in entertainment demand: Evidence from Wimbledon

July 12, 2016

Abstract

This paper empirically examines how suspense and surprise affect the demand for entertainment. We use a tennis tournament, the Wimbledon Championships, as a natural laboratory. This setting allows us to both operationalize suspense and surprise by using the audience's beliefs regarding the outcome of the match and observe the demand for live entertainment using TV audience figures. Our match fixed effects estimates of 8,563 minute-by-minute observations from 80 men's singles matches between 2009 and 2014 show that both suspense and surprise are drivers of media entertainment demand. In general, surprise seems to be more important in this regard than suspense, and both factors matter more during a match's later moments. We discuss important implications for the design of entertainment content to maximize entertainment demand.

JEL Classification: D83, L82, L83

Keywords: Suspense, Surprise, Entertainment, TV audience, Betting odds, Tennis

1 Introduction

Media entertainment plays an important role in people's daily lives. Vorderer, Klimmt, and Ritterfeld (2004) describe media entertainment as enjoyment from consuming media content, whether at home or at an outside venue. Given that entertainment providers are facing stiffer competition in the entertainment market, understanding precisely what factors drive the demand for entertainment content is of critical importance.

Previous studies have identified *suspense* and *surprise* as two major determinants of enjoyment associated with media consumption (e.g., Zillmann, 1991, 1996; Vorderer et al., 2004). The online Cambridge English Dictionary defines suspense as "a feeling of excitement or anxiety while waiting for something uncertain to happen" and surprise as "an unexpected event, or the feeling caused when something unexpected happens." Importantly, both occur exclusively in situations in which there is concern over uncertain outcomes (Comisky & Bryant, 1982).

Suspense and surprise are best understood and modeled in a Bayesian setting (Ely et al., 2015). In this setting, probabilities quantify personal beliefs: people form hypotheses about the occurrence of specific events (e.g., "it will rain tomorrow") and attach probabilities to them based on their subjective levels of belief in these hypotheses ("with a 90% probability"). In the Bayesian view, people will transform their *prior* beliefs into *posterior* beliefs when new and relevant information arrives (Itti & Baldi, 2009). This continuous process of forming and updating beliefs leads to entertainment based on the experience of suspense and surprise, where suspense and surprise are the forward- and backward-looking emotions, respectively.

Suspense evolves through the assessment of future events, with a moment carrying more suspense when some crucial uncertainty is soon to be resolved (Vorderer et al., 2013), such as a researcher opening a letter with the committee's decision on his or her research grant application. By contrast, surprise evolves by assessing past events, with a moment carrying more surprise immediately after an unexpected event occurs (Itti & Baldi, 2009), such as after an underdog soccer team scoring the winning goal.

Although it is intuitive that suspense and surprise matter in the context of entertainment, empirical tests are difficult to design because people's beliefs and their enjoyment are hard to observe. Moreover, little is known about the importance of suspense relative to surprise or about their importance with respect to the passage of time. In this paper, we address these questions by employing high-frequency data from a tennis tournament, the Wimbledon Championships, which offers our research two unique advantages.

The first advantage is that we can quantify the audience's beliefs because modeling tennis situations is possible. In tennis, a Bayesian audience forms beliefs about the final outcome of the match, i.e., about the likelihood that a particular player will win a particular match.[1] We estimate the relevant beliefs at the point-by-point level in two ways: first, we use a Markov model that requires the player's probability of winning a service point and the current score as inputs; second, we use in-play betting odds.

The second advantage that tennis offers is that the demand for entertainment is observed using high-frequency minute-by-minute live TV audience figures (*ratings*) during the matches. As viewers can easily – and at no cost – switch channels or turn off the TV to maximize their utility from viewing, short-term variations in TV audience figures reflect whether the audience is enjoying a given match (Alavy, Gaskell, Leach, & Szymanski, 2010). By using minute-by-minute information regarding aggregate viewers' behavior, we can uncover an audience's underlying preferences for entertainment in a real-world environment.[2]

This paper contributes to the literature by presenting an analysis of unique and naturally occurring field data that provide a rare opportunity to empirically investigate the importance of suspense and surprise when consuming a media entertainment product. Our empirical analyses reveal that both suspense and surprise have a positive effect on entertainment demand. Using 8,563 minute-by-minute observations from 80 men's singles matches between 2009 and 2014, our match fixed effects estimates reveal that minutes with more surprise and suspense have significantly higher live TV ratings. This result indicates that suspense and surprise are complementary and that demand for entertainment is stronger for higher levels of suspense and surprise. In particular, a one standard deviation increase in suspense (surprise) is associated with an audience increase of approximately 1,200 (2,200) viewers per minute. For some perspective,

[1] The same idea can be applied in other settings. For example, people assign probabilities to the hypothesis that a president will be reelected, that a mission will succeed, or that a company's earnings will beat analysts' consensus estimates. What differentiates tennis from other settings is the frequency with which events happen and new information is revealed.

[2] TV remains the central provider of entertainment content despite the increasing supply of entertainment available on the Internet. According to the U.S. Bureau of Labor Statistics, individuals aged 15 and over watched TV for 2.8 hours per day on average in 2013, accounting for more than half of their leisure time.

the minute-level effect of a one standard deviation increase in suspense and surprise combined corresponds roughly to a 3% audience increase (based on an average audience of approximately 100,000 viewers in our sample). Although we cannot compare our results with those from previous studies, our estimates suggest that the impact of suspense and surprise on TV audience figures is economically non-trivial.

Moreover, we find that the audience impact of surprise is consistently greater than that for suspense: depending on the model used for computing the audience's beliefs, the estimated effects for surprise are between two and five times greater than those for suspense. Hence, surprise appears to be more important than suspense in entertainment demand. In addition, over the course of a match, the impact of both suspense and surprise clearly increases. This implies that the entertainment effect of suspense and surprise is larger when the stakes are higher.

To the best of our knowledge, this paper is the first to test Bayesian theory on suspense jointly with surprise under natural conditions. We provide a framework that entertainment industry managers can use to measure an audience's beliefs, which can then be used to measure entertainment from suspense and surprise. Although in tennis there is not much room for artificially increasing suspense and surprise, the implications of our study are far more important for other entertainment settings in which content can be designed ad hoc to increase the public's enjoyment. Designers of films, TV series, TV shows, online videos, novels, or gambling games should be aware of people's preferences for suspense and surprise, their increasing significance towards the end of a media event, and the greater importance of surprise.

The remainder of this paper is organized as follows. Section 2 reviews the literature. Section 3 describes our setting and data. Section 4 outlines the operationalization of suspense and surprise and our empirical methodology. Section 5 presents the empirical results and various robustness checks. Section 6 discusses several implications and concludes.

2 Literature review

The theoretical literature on suspense and surprise is limited. Yet Ely et al. (2015) recently filled this gap by introducing a framework in which a Bayesian audience derives entertainment utility (enjoyment) from *anticipated* changes in beliefs (suspense) and *actual* changes in beliefs (surprise). In the model developed by these authors, higher suspense results from greater variance in the next period's beliefs – what is currently happening versus what is expected to happen next – and higher surprise results from greater distance between previous and current beliefs.

However, no study has yet empirically investigated the relationship between suspense, surprise and enjoyment. Most of the relevant studies are laboratory experiments that focus either on suspense or surprise. For example, Bryant et al. (1994) recorded a football match and manipulated its commentary to create a high-suspense version and a low-suspense version. After having watched one of these two versions, participants were given a questionnaire and asked to rate their enjoyment on a scale from 0 to 10. The results from Bryant et al. (1994) show that viewing the high-suspense version was significantly more enjoyable and exciting. Su-lin et al. (1997) and Peterson and Raney (2008), using students as participants, examine suspense as a factor in the enjoyment of National Collegiate Athletic Association (NCAA) men's basketball games. Both studies operationalize suspense as the final point differential in a game and enjoyment as the average of seven different items (e.g., "the game excited me" or "I enjoyed the game") that are rated on a scale from 0 to 10. Their results show that higher suspense leads to greater enjoyment.

In an experimental setting, Itti and Baldi (2009) test whether surprise attracts the attention of participants when watching one of several videoclips, including television broadcasts, such as news, sports, commercials, and outdoor scenes. They define surprise in Bayesian terms as the distance between the prior and posterior distributions of beliefs, and they employ eye-tracking technology to measure attention. Their results demonstrate that surprise explains the greatest portion of human eye movements, indicating that humans are attracted to surprising elements in video displays.[3] Moreover, in a laboratory setting, Alwitt (2002) finds that viewers perceive

[3] Baldi and Itti (2010) build on these results and provide further evidence of this relationship. For a recent review of studies on the surprise-attention link, see Horstmann (2015).

suspenseful TV commercials as shorter, attributing this effect to the viewers' intensified attention and interest.

In the sports economics literature, in particular, suspense has been often associated with "uncertainty of outcome." The uncertainty of outcome hypothesis posits that more uncertainty about the final winner of a sports competition leads to more suspense (Borland & MacDonald, 2003). Alavy et al. (2010) use minute-by-minute audience figures from 248 English Premier League matches to measure the effect of outcome uncertainty on entertainment demand. They find that matches with less probability of ending in a draw – but also with less score differential between teams – generate more viewers. In a sports-related article, Olson and Stone (2014) model viewers' entertainment as a function of suspense to evaluate whether the introduction of post-season playoffs in U.S. college football would be an improvement over the current system. Using match-level Nielsen ratings over 2011-2013 for 70 football and basketball matches, they show that the level of viewers' entertainment significantly increases with the championship's suspense level.

3 Setting and data

Clearly, a tennis match provides many moments of various levels of suspense and surprise: comebacks, break points, tie-breaks, injuries, and spectacular rallies can make any moment entertaining, whereas unimportant points can make any moment less entertaining to watch.[4] We use the 2009 Wimbledon men's final between Andy Roddick and Roger Federer as our illustrative example. After winning the first set, Roddick had four set points in the second set, putting him only one set away from the championship. However, supported by his strong service, Federer won all of Roddick's set points and eventually won the set. Because the audience had to strongly readjust its beliefs about Federer's chances of winning, we describe such circumstances as surprising. In the final set, which is played until one player wins at least six games by at least a two-game spread, Federer finally won 16-14 in a set that lasted more than 90 minutes. Because each point could potentially bring a player's winning probability very close to either zero or one, we describe such circumstances as suspenseful.

Our definition of suspense and surprise is best understood in a Bayesian framework. Simply

[4] In Appendix A, Part I briefly describes the rules of tennis and its jargon.

put, a Bayesian audience has some current beliefs, based on the information currently available, about a certain outcome. Upon the arrival of new, relevant information, the audience will update its beliefs, which are then called posterior beliefs. In tennis, viewers form and continuously update their beliefs about the "hypothesis" that a given player might win the match. To quantify moments of various levels of suspense and surprise, we must therefore estimate the relevant beliefs at the point-by-point level. We do this using two methods.

In the Markov chain method, we rely on the explicit structure of the data-generating process in tennis. In tennis, points are linked to games, games to sets, and sets to matches; thus, a match can be modeled as a binary Markov chain (Newton & Aslam, 2009; O'Malley, 2008).[5] We estimate the unique belief path for each match using a computer program that computes the likelihood of winning the match for a given player point-by-point over the match.[6] The only input for this simple model is the score and the probability of winning a point on serve. Detailed match data at the point level are provided by IBM, the official supplier of information technology to the Wimbledon Championships. Beyond general information about the match, such as the players, courts, start and end match times, these data also contain point-by-point information on the current score, time (exact to the second), server, and winner.

In the betting odds method, we rely on the information content of in-play betting odds from *in-play* betting markets. Wolfers and Zitzewitz (2006) show that betting odds provide valuable estimates of average, aggregate beliefs about the probability that an event will occur. The odds originate from Betfair, one of the largest online betting markets, and are provided by Fracsoft, a data vendor.[7] Betfair's online platform provides a market for opinions and for participants to bet against one another by offering and accepting odds under which they are willing to buy or sell a certain bet. Bettors mostly follow the match live on TV or on other electronic devices and continuously place their bets during the match: whenever new information becomes observable, they update their beliefs, and the odds change accordingly.

[5] Walker et al. (2011, p.490) illustrate the binary Markov scoring rule for a game of tennis. Moreover, Liu (2001) and Barnett and Clarke (2005) propose similar ways of modeling the probability of winning a match.

[6] The software program, described in Klaassen and Magnus (2014), is called "Richard" and is freely available online at www.janmagnus.nl/misc/wimbledon.pdf with detailed instructions.

[7] Trading volumes on Betfair are very large. For example, 1.2 billion bets were placed in 2014, resulting in a total trading value of roughly \$92 billion. Croxson and Reade (2014) estimate that the daily trading intensity on Betfair during the 2005-2007 period was greater than the daily trading intensity on all the major European Stock Exchanges combined. With regard to our sample, we observe an average total trading volume of \$28.5 million per match, 70% of which was placed in-play. Approximately \$78 million worth of bets were placed on the 2014 Wimbledon men's final alone, and 92% of that amount was placed in-play.

Importantly, as the Markov belief is based on the actual score and on the server's probability of winning a point, the in-play betting odds belief should be more accurate. In fact, odds would reflect not only the newest information, such as a player's injury, but also a number of historical factors, such as a player's performance record on grass courts or previous head-to-head records between the players. In this sense, we consider the betting odds method as the main specification.[8] The odds enable us to derive the aggregate market's belief at each point in the match about a given player's probability of winning the match. Indeed, the inverse of the odds on the expected match winner can be interpreted as the aggregate current belief about a player's match winning probability (Hasbrouck, 1991).

To measure entertainment demand, we gather high-frequency TV audience ratings over the 2009-2014 period on all Wimbledon men's singles matches that were transmitted live on the Swiss national German-language channels *Schweizer Fernsehen Zwei* (SRF2) and *Schweizer Fernsehen Info* (SRFinfo), two of the largest Swiss broadcaster's free channels.[9] In Switzerland, Wimbledon – and tennis as a whole – enjoys good TV coverage. Overall, SRF broadcasted 108 Wimbledon men's singles matches between 2009 and 2014. In our analysis, we exclude 28 matches: three matches are excluded because no betting data is available, whereas 25 partially and shortly transmitted matches are excluded because of potential spillover effects in the audience ratings caused by the preceding and following TV programs.[10]

Mediapulse, a Swiss ratings firm essentially equivalent to Nielsen, generates audience statistics using survey data from a panel encompassing 1,870 households across Switzerland, which contains approximately 4,200 people three years of age and over. As the advertising industry also uses Mediapulse's data, the panel must meet strict requirements to reflect the Swiss population as accurately and representatively as possible (see Appendix B for further details). TV audience ratings measure the total number of single viewers watching the channel at each moment. For

[8]We thank an anonymous referee for this suggestion.

[9]Whereas SRF2 focuses on either live or recorded sports programming, SRFinfo chiefly rebroadcasts programs from SRF1 (the first national channel) and SRF2. However, it also occasionally acts as a complementary channel for live sportscasts in the event of programming conflicts. On average, in our sample, SRF2 has 112,515 viewers, whereas SRFinfo has 45,853 viewers.

[10]Although the average duration of the excluded matches is only 14 minutes, one might worry that the exclusion of 25 matches introduces a selection bias because the broadcaster might stop broadcasting matches with low suspense, low surprise, or both. As we cannot compute the levels of suspense and surprise due to IBM data unavailability, we examine the TV channel content description (the list is available upon request) and also directly asked the broadcaster. We discovered that the broadcaster cannot observe live audience figures: as a result, programming decisions do not depend on such live audience figures. It rather appears that the broadcaster "fills" programming "gaps" with some scenes from these 25 matches.

example, a rating of 122,500 indicates that 122,500 single viewers were tuned into the program on average during a particular minute. For any minute when more people tune into the tennis match (either from another channel or by turning on the TV) than turn off the match, the rating will increase.

Our final data set consists of roughly 8,500 minute-by-minute observations on detailed match statistics, in-play betting odds, and live TV ratings. In Table 1, Panel A reports descriptive statistics for the dependent variable *audience* on the 80 matches in the final sample. An average match has an audience of slightly more than 106,000 spectators, corresponding to a market share of 16.9%. Panel B additionally shows that *audience* is larger at later tournament stages. Our sample is heterogeneous, containing matches from the first stage up to the finals and a total of 58 unique players. An average match lasts 107 minutes, consists of 120 points and 3.6 sets, and each set consists of 10 games. Late tournament matches are longer, as they typically are more balanced.

Table 1
Descriptive statistics of the audience and sample characteristics.

Panel A: Descriptive statistics

Variable	Description	N	Mean	Std. Dev.	Min	Max
audience	TV Rating	8,563	106.47	136.81	1.77	1,083.7

Panel B: Sample characteristics

Tournament stage	N	Length (minutes)	Points in match	Sets in match	Games in set	*audience*
1*st* stage	12	85	101	3.25	9.41	44.0
2*nd* stage	14	83	102	3.28	9.61	55.2
3*rd* stage	12	98	121	3.75	9.44	75.8
4*th* stage	10	113	118	3.67	9.94	83.2
Quarterfinal	14	102	128	3.86	10.23	87.4
Semifinal	12	138	127	3.75	10.57	89.1
Final	6	162	150	4	11.33	362.8
Full sample	80	107	118.6	3.62	10.01	106.47

Notes: The table reports descriptive statistics for the 80 men's singles matches played at Wimbledon between 2009 and 2014. Panel A describes the TV audience ratings (in thousands, except for N), whereas Panel B provides the means of additional match characteristics (by tournament stage and for the full sample).

4 Estimation approach

4.1 Suspense and surprise

The Markov method

The construction of suspense and surprise closely follows the work of Ely et al. (2015), in which the entertainment utility of the Bayesian audience is a function of the Markov belief path. First, we model suspense in the form of an expectation, where higher suspense is attributed to greater variance in the next period's belief.[11] For the point p:

$$SUS_p^{Markov} = [\mathrm{E}\sum_{\omega}(\mu_{p+1}^{\omega} - \mu_p)^2]^{1/2}, \tag{1}$$

where μ_p refers to the current player's probability of winning the match (the current belief) at the moment when point p is scored and μ_{p+1}^{ω} refers to the anticipated posterior probability of a player's winning the match. The posterior belief depends on the realization of state ω, which in this setting is a binary variable: $\omega = 1$ when the server i wins the point (with probability S_i) or $\omega = 0$ when he loses it (with probability $1 - S_i$). Thus, we can rewrite equation (1) as follows:

$$SUS_p^{Markov} = [S_i \cdot (\mathrm{E}[\mu_{p+1}^{1}] - \mu_p)^2 + (1 - S_i) \cdot (\mathrm{E}[\mu_{p+1}^{0}] - \mu_p)^2]^{1/2}. \tag{2}$$

Because only one state occurs in reality, we also estimate the *counterfactual* posterior belief, defined as the probability of winning the match for the unobserved state.[12] To do so, we replace the actual score with the counterfactual score in the Markov model. For example, if player i actually serves and wins the first point of the match ($\omega = 1$), the counterfactual belief is computed by assuming that he *lost* the first point. The player's probability of winning a service point (S_i) is computed for each player in our sample based on historical data from the Association of Tennis Professionals (ATP) website.[13] To obtain accurate predictions from the model, we set μ_0, i.e., the winning probability at the beginning of the match ($p = 0$), equal to the corresponding odds-implied probability.[14]

[11] In Appendix C, Part I illustrates how we compute SUS_p^{Markov} and SUR_p^{Markov} with a numerical example.
[12] According to equation (2), if the server actually wins (loses) the next service point, the counterfactual is represented by $\mathrm{E}[\mu_{p+1}^{0}]$ ($\mathrm{E}[\mu_{p+1}^{1}]$).
[13] In Appendix A, Part II provides further details regarding how we compute S_i.
[14] We are grateful to an anonymous referee for this suggestion.

Second, we model surprise in the form of the Euclidean distance between the prior belief and the current belief, where greater surprise results from the occurrence of an event that strongly contradicts the audience's belief, constraining the audience to change its beliefs (Itti & Baldi, 2009). For the point p:

$$SUR_p^{Markov} = |\mu_p - \mu_{p-1}|, \tag{3}$$

where μ_{p-1} refers to the probability of winning the match for a player one point earlier.

The betting odds method

The construction of suspense and surprise based on betting odds is straightforward and similar to the Markov method. Following Hasbrouck (1991), we compute the average mid-odds from the best buy $odds_{ip}^{back}$ and sell $odds_{ip}^{lay}$ for player i for each point p in match m as follows:

$$odds_{ip}^{mid} = \frac{odds_{ip}^{back} + odds_{ip}^{lay}}{2} \quad \forall(p, m), \tag{4}$$

from which we derive the implied winning probability for player i:

$$\tilde{\nu}_{ip} = \frac{1}{odds_{ip}^{mid}} \quad \forall(p, m). \tag{5}$$

Although the sum of the winning market probability for player i and player j ($\tilde{\nu}_{ip} + \tilde{\nu}_{jp}$) should sum up to one in a frictionless market, in practice it rarely does so because of transaction costs. Following the standard approach to eliminating this overround (e.g., Forrest et al., 2005), we adjust the implied winning probability to obtain the final market implied winning probability for player i as follows:

$$\nu_{ip} = \frac{\tilde{\nu}_{ip}}{\tilde{\nu}_{ip} + \tilde{\nu}_{jp}} \quad \forall(p, m). \tag{6}$$

As the change in the implied winning probability is symmetric, the choice of which player's probability of winning to consider is irrelevant, so we can drop the player subscript i.

First, we define surprise for point p as follows:

$$SUR_p^{odds} = |\nu_p - \nu_{p-1}|, \tag{7}$$

where ν_p refers to the current player's probability of winning the match (the odds-implied current belief) at point p and ν_{p-1} refers to the odds-implied prior belief at the point $p-1$. The only difference from equation (3) involves the type of data used.

Second, as we cannot implement the baseline metric of suspense by relying entirely on betting odds, we redefine the baseline suspense measure, but only slightly. By definition, betting odds reflect bettors' *current* beliefs at any point in the match, based on a given information set. However, we do not know which value the odds would have taken had the next point gone differently, i.e., the counterfactual posterior belief. Therefore, we come up with a "hybrid" implementation of the baseline suspense measure in which we use the Markov chain model only to determine the counterfactual posterior belief ($\mathrm{E}[\mu_{p+1}^{\omega}]$) for the unobserved state. Because the audience must also estimate the counterfactual probability, this procedure is suitable for our purposes. Analogously to equation (2), we model suspense for point p as follows:

$$SUS_p^{odds} = \begin{cases} [S_i \cdot (\nu_{p+1} - \nu_p)^2 + (1 - S_i) \cdot (\mathrm{E}[\mu_{p+1}^0] - \nu_p)^2]^{1/2} & \text{if } \omega = 1, \\ [S_i \cdot (\mathrm{E}[\mu_{p+1}^1] - \nu_p)^2 + (1 - S_i) \cdot (\nu_{p+1} - \nu_p)^2]^{1/2} & \text{if } \omega = 0. \end{cases} \tag{8}$$

Against the backdrop of equation (2), we now substitute the Markov current belief μ_p with the odds-implied current belief ν_p. Concerning the posterior beliefs, we must distinguish between the states ω: when the server wins the service point ($\omega = 1$), we substitute $\mathrm{E}[\mu_{p+1}^1]$ with the actual odds-implied posterior belief ν_{p+1}, and we use the Markov posterior belief for the counterfactual ($\mathrm{E}[\mu_{p+1}^0]$); when the server loses the service point ($\omega = 0$), we substitute $\mathrm{E}[\mu_{p+1}^0]$ with the actual odds-implied posterior belief ν_{p+1}, while we use the Markov posterior belief for the counterfactual ($\mathrm{E}[\mu_{p+1}^1]$).

Finally, because we observe minute-by-minute variation in the TV audience, we translate suspense and surprise from point level to minute level by computing the average suspense or surprise over the minute during which more than one point is scored within a minute.[15]

[15] We also test an alternative method that considers only the last point scored in any minute when more than one point is scored. This alternative does not alter our results.

A Comparison of the Markov and the betting odds beliefs

Figure 1 shows the belief paths regarding Novak Djokovic's chance to win the match he played against Roger Federer on 7th June, 2014. The black line shows the belief path computed using the betting odds, whereas the grey line shows the Markov belief path. Both beliefs start at 60%, the implied probability of winning from the in-play betting odds – apparently, the bettors thought Djokovic was the favorite – and end at 100% for Djokovic, who won the final in five sets.

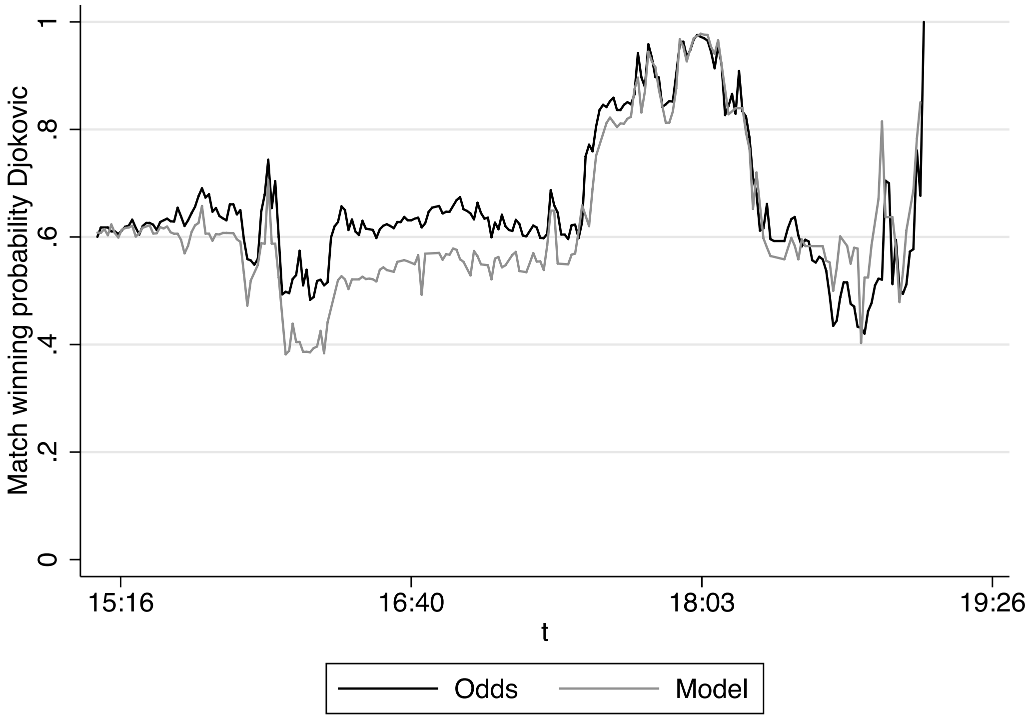


Figure 1 Displayed are the belief paths about Novak Djokovic's chance to win his 7th June, 2014 match against Roger Federer. Djokovic won in five sets with the score: 6-7; 6-4; 7-6; 5-7; 6-4. A total of $72 million was bet *in-play* on this Wimbledon final.

Notably, the belief path from the Markov model highly correlates (0.991) with the belief path extracted from the betting odds. Allegedly, professional bettors also strongly rely on computer programs to estimate the underlying player's winning probability during a match (Hutchins, 2014). Nonetheless, the Markov and betting odds belief paths might slightly diverge due to the arrival of new, relevant information beyond server and score, such as the signs of an injury, weather changes, or the stadium atmosphere.[16] However, because suspense and surprise are computed from *changes* in beliefs, the size of the vertical gap between the odds belief and the Markov belief is irrelevant – it is much more important that the beliefs are positively correlated.

Table 2 reports summary statistics for the suspense and surprise variables. Means and standard deviations of suspense and surprise are comparable between models, although both are slightly higher when computed with the odds-derived belief. The correlation between suspense and surprise is 0.36 (0.38) for the Markov chain (betting odds) method. Finally, we observe that suspense and surprise increase over the stages of the tournament.[17]

Table 2
Descriptive statistics of suspense and surprise.

Variable	Description	N	Mean	Std. Dev.	Min	Max
SUS^{Markov}	Suspense based on Markov	8,563	0.0149	0.0410	1.93e-07	0.2121
SUR^{Markov}	Surprise based on Markov	8,563	0.0122	0.0179	0	0.2717
SUS^{odds}	Suspense based on betting odds	8,563	0.0233	0.0256	4.99e-07	0.2601
SUR^{odds}	Surprise based on betting odds	8,563	0.0129	0.0197	0	0.3097

Notes: The table reports summary statistics for the 80 men's singles matches played at Wimbledon between 2009 and 2014, for a total of 8,563 minutes of live tennis.

4.2 Empirical methodology

The type of data used in this study presents two advantages with respect to the estimation methods. First, our panel data allow us to control for time-invariant factors that might jointly affect the audience level by using within-match variation. Those factors might be the stage of the competition, the day of the week, or the quality of the players.[18] Second, as opposed to stadium attendance, short-term TV audience variation is not affected by factors such as supply

[16] Another cause might rely on match-fixing. In January 2016, BBC and BuzzFeed News uncovered evidence of widespread suspected match-fixing in tennis, including some matches at Wimbledon (see, for example, the article of Simon Cox at www.bbc.co.uk/sport/tennis/35319202). Because match-fixing could bias the betting odds, it could also bias the odds-derived belief of a player's winning probability. However, we believe that our results are not systematically affected for two main reasons: first, the report refers to events from approximately ten years earlier (two matches listed in the report are from 27 June 2006 and 26 June 2007), i.e., a period not covered by our sample, and second, although some of our matches may have been the subject of match-fixing without our knowledge, our suspense and surprise measures are based on *changes* in the winning probability, not on absolute levels.

[17] Descriptive statistics of the main variables at the tournament-stage level are provided in Appendix C, Part II.

[18] Rodríguez et al. (2015) illustrate the importance of including a large set of control variables when examining aggregate TV audience measures. As their dependent variable is the average TV audience over the length of the program (they analyze professional cycling races), they also control for, among other things, calendar variables, the scheduling of rival channels, and the competitive balance before the race.

capacity, gate price, location, or weather conditions (Borland & MacDonald, 2003; Alavy et al., 2010).

Simultaneously, to account for other characteristics that are subject to change *during* a match that might affect the minute-by-minute TV ratings, we use several control variables. To begin with, we introduce time dummies that correspond to the elapsed time (in minutes) from the start of the match. This allows us to control for any time-related audience differences in a flexible manner. Because all channels offer their best TV content in the evening hours, potentially turning viewers away from tennis, we introduce a *primetime* dummy, which takes the value of one for all the minutes after 8:00 p.m. and zero otherwise.[19] Intuitively, the coefficient for *primetime* should be negative.

Due to the popular news program on SRF1 (the first national channel), the TV audience level of SRF2 might drop as viewers switch to the newscast. Therefore, we construct the *news* indicator variable, which equals one for any minute between 5:58–6:06 p.m. and between 7:28–7:56 p.m., indicating the first short newscast (6:00–6:05 p.m.) and the following long one (7:30–7:55 p.m.), respectively, and zero otherwise. Intuitively, the coefficient for *news* should be negative. Other programs are not likely to systematically affect the audience variation, particularly because the matches take place at various times of the day, from Monday through Sunday. As Alavy et al. (2010) note, minor variations in audience might be due to channel hoppers. Nonetheless, these authors argue that any moment providing high entertainment should attract even channel hoppers and keep them watching, thus reducing the noise from their behavior.

To allow the players to rest and switch sides of the court, small breaks take place after odd-numbered games and between sets. As these breaks may cause viewers to briefly stop watching, we introduce the *pause* indicator variable, which equals one during the break and zero otherwise.[20] The *pause* variable is needed only for the regression based on betting odds. By construction, the suspense and surprise measures computed with the Markov model are missing during the break because the score does not change.

[19] The results are robust to several alternative definitions of *primetime*, e.g., 6:00 p.m. and 7:00 p.m.

[20] TV broadcasters also use certain breaks for showing commercials. However, we do not insert an *advertising* dummy because the *pause* dummy also captures the effect of viewers switching channels to skip the advertising. To identify time-outs for advertising, we examine the channel content description for SRF2 and SRFinfo, which lists all the programs broadcast and their exact start and end times. Commercial breaks occur on average on a 35-minute basis for all non-finals matches and on a 20-minute basis for the finals. As SRF2 is a national public channel, the amount of advertising is limited by law.

Regarding the estimation model, as the Wooldridge test evidenced autocorrelation in the audience ratings, we add four lags of *audience* into the regression equation (Wooldridge, 2010). We therefore estimate a dynamic model by applying the Arellano-Bond generalized method of moments (GMM) technique, as the GMM estimators are consistent estimators for dynamic panels (Arellano & Bover, 1995).

Our baseline regression model is specified as follows:

$$\begin{aligned} audience_{i,t} = \alpha_0 &+ \beta_1 \cdot SUS_{i,t} + \beta_2 \cdot SUR_{i,t} \\ &+ \beta_3 \cdot primetime_{i,t} + \beta_4 \cdot news_{i,t} + \beta_5 \cdot pause_{i,t} \\ &+ \theta_k \cdot \sum_{k=1}^{4} audience_{i,t-k} + time\ dummies + \upsilon_i + u_{i,t} \ , \end{aligned} \tag{9}$$

where the subscripts i and t denote match and minute, respectively, and υ_i denote the match fixed effects. The dependent variable $audience_{i,t}$ represents the match i's average live TV audience level at minute t, $\sum_{k=1}^{4} audience_{i,t-k}$ the four lags of the dependent variable, and *time dummies* represent the minutes elapsed since the beginning of the match. The coefficients of interest are β_1 for suspense (SUS) and β_2 for surprise (SUR). For all regressions, we use robust standard errors that are clustered at the match level. Diagnostic and robustness checks are discussed in Subsections 5.2 and 5.3.

5 Empirical results

5.1 Univariate evidence

Before we turn to the estimation results of equation (9), we report the results of a univariate analysis of our data. Due to programming conflicts, the broadcasting of tennis is sometimes switched from SRFinfo to SRF2 or vice versa. These switches occur independently of the standing of the current match and thus independently of the moment's level of suspense and surprise. Figure 2 depicts such an example from a 2012 third-round match between Roger Federer and Julien Benneteau, initially broadcast on SRFinfo and then switched to SRF2 at 9:35 p.m.

Clearly, the audience switched channels to follow the match, indicating that the audience was actively following. However, the relevant question is whether the size of the audience change depends on the levels of suspense and surprise during the previous minutes. To answer this question, we identify eight switches that occurred during the broadcast of matches in our sample. For each switch, we can define the precise moment when the live broadcast was interrupted and then continued on the other channel. We measure the jump in ratings on the channel where the broadcast is continued by taking the difference between the average 15-minute ratings preceding and following the switch. We sum SUS^{Markov} and SUR^{Markov} over 15 minutes before the switch, sum the two totals, and compute the median. We create two groups based on whether suspense and surprise were above or below the median. Intuitively, if the 15 minutes before the switch offered above-median (below-median) entertainment, this should be reflected in a larger (smaller) jump in the TV audience on the post-switch channel.

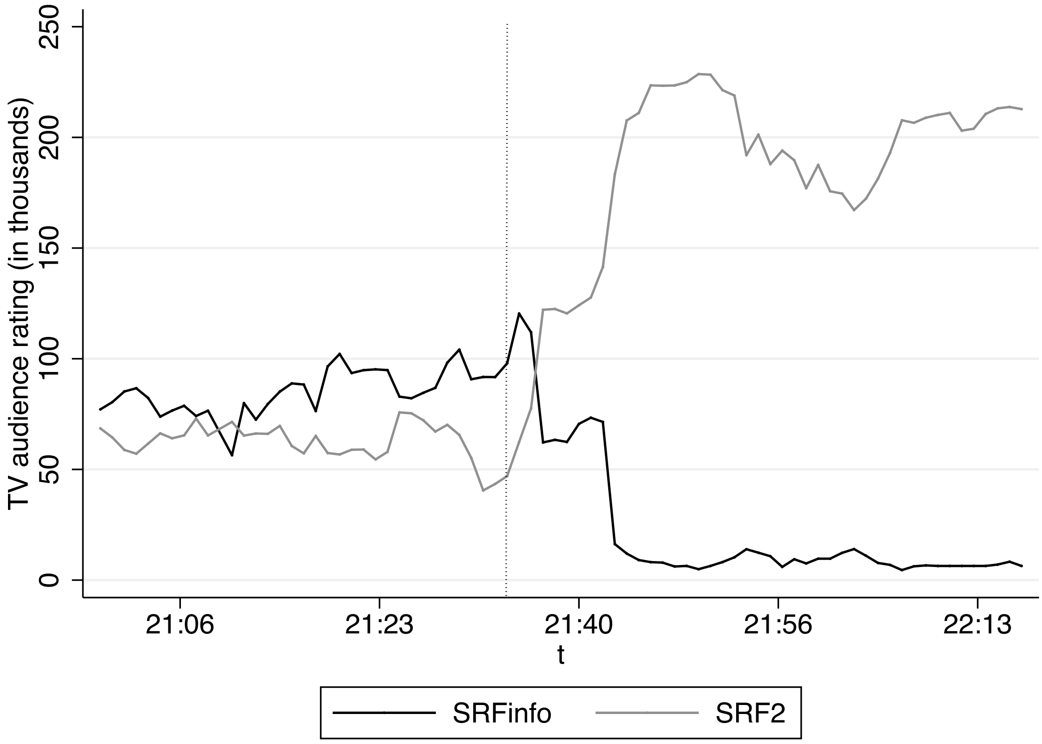


Figure 2 Displayed is the evolution of the TV audience ratings around an exogenous channel switch. The match between Federer and Benneteau was initially broadcasted on SRFinfo and then switched to SRF2 at 9:35 p.m, as indicated by the vertical dotted line. For about three minutes, both channels showed the same content. The SRF2 and the SRFinfo are two free Swiss national channels.

We use the Wilcoxon rank sum test to compare differences in TV audience variation after a broadcaster-initiated channel switch between groups with low and high suspense and surprise. We find that the group with below-median suspense and surprise shows an increase in TV audience of 29,900 viewers, while the group with above-median suspense and surprise shows an increase in TV audience of 99,450 viewers ($z = 2.309$, $p < 0.05$).[21] Overall, this finding provides not only evidence for an active audience assumption but also preliminary and suggestive evidence supporting the hypothesis that a TV program's level of entertainment, measured in terms of suspense and surprise, affects its performance in terms of audience.

5.2 Regression analysis

Table 3 reports regression estimates for the effects of suspense and surprise on TV audience ratings. The Arellano-Bond tests for autoregressive errors yield the expected results (Arellano & Bover, 1995): autocorrelation exists in the first lag but not in the second. Additionally, the Sargan tests of over-identifying restrictions support the null hypothesis that the instruments are valid.

Overall, the results are in line with the hypothesis that moments that offer more suspense and surprise generate more entertainment demand. Columns (1)-(3) show the results for suspense and surprise measures based on the Markov chain model. Columns (1) and (2) demonstrate that suspense and surprise each have a positive effect on TV ratings. Column (3) shows that even when we include both variables together, suspense and surprise remain significant predictors for the TV audience. This result indicates that suspense and surprise are both driving forces behind media entertainment demand. On average, a one standard deviation increase in suspense raises the audience by approximately 1,260 viewers per minute, whereas a one standard deviation increase in surprise raises the audience by approximately 2,630 viewers per minute. As an illustration, the combined effect of a one standard deviation increase in both suspense *and* surprise results in an increase of approximately 3,900 viewers per minute, corresponding to a 3.65% minute-level increase (with respect to an average audience per match of 106,000).

Columns (4)-(6) present the results for suspense and surprise measures based on the in-play betting odds. In all specifications, suspense and surprise have a positive and significant coefficient. According to column (6), on average, a one standard deviation increase in suspense raises the audience by approximately 1,140 viewers per minute, whereas a one standard deviation increase in surprise raises the audience by approximately 1,880 viewers per minute. As an

[21] The results are similar when using suspense and surprise from the betting odds method.

Table 3
The effect of suspense and surprise on TV audience.

	Dependent variable: *audience*					
	Markov			Betting odds		
	(1)	(2)	(3)	(4)	(5)	(6)
SUS	46.542***		30.803***	56.867***		44.512***
	(12.883)		(10.122)	(16.824)		(15.472)
SUR		172.861***	146.984***		144.051***	95.404***
		(28.570)	(29.401)		(30.355)	(31.081)
primetime	0.463	0.870	1.064	0.871	0.904	1.167
	(0.412)	(0.687)	(0.660)	(0.631)	(0.751)	(0.770)
news	-1.692***	-1.511***	-1.448**	-1.719***	-1.775***	-1.603**
	(0.531)	(0.590)	(0.614)	(0.614)	(0.659)	(0.665)
pause				-4.138***	-3.761***	-3.918***
				(0.842)	(0.804)	(0.799)
Audience lags	Yes	Yes	Yes	Yes	Yes	Yes
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes
Match fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
N	8,332	8,332	8,332	8,387	8,387	8,387
AR(1) z test	-4.59	-4.62	-4.62	-4.44	-4.54	-4.49
$Pr > z$	0.000	0.000	0.000	0.000	0.000	0.000
AR(2) z test	0.35	0.11	-0.05	-0.17	-0.34	0.03
$Pr > z$	0.724	0.910	0.958	0.863	0.734	0.972
Sargan test $Pr > \chi^2$	1.000	1.000	1.000	1.000	1.000	1.000

Notes: The table reports the results of panel regressions using the Arellano-Bond GMM estimation method. The dependent variable is the TV audience rating (in thousands). The main independent variables, suspense and surprise, are derived from the Markov model (Markov) and in-play betting odds (Betting odds). All estimations also include a constant (not reported). Time dummies correspond to the elapsed time (in minutes) from the start of the match. The data are at the minute-level and includes 80 men's singles matches played at Wimbledon that were broadcast live on SRF2 and SRFinfo from 2009 to 2014. Robust standard errors that have been adjusted for clustering at the match level are provided in parentheses. In all models, *, **, and *** denote significance at the 10%, 5% and 1% levels, respectively.

illustration, the combined effect of a one standard deviation increase in suspense *and* surprise results in an increase of approximately 3,000 viewers per minute, corresponding to a 2.83% minute-level increase (with respect to an average match audience of 106,000). Thus, the economic effects estimated with both methods are comparable.

Moreover, surprise appears to be more important than suspense in entertainment demand. The t-statistic of the equality of the estimated coefficients for suspense and surprise is strongly significant for the Markov method ($\chi^2 = 12.39$, $Pr > \chi^2 = 0.000$) and marginally insignificant for the betting odds ($\chi^2 = 1.76$, $Pr > \chi^2 = 0.184$). Depending on the model used to estimate the audience's beliefs, the estimated coefficients for surprise are from two to five times larger than those for suspense.

The coefficients of the control variables mostly have the signs that are expected. The coefficient for the first lag of audience is always very close to one, suggesting some short-term inertia in viewership, whereas the other three lags are small in magnitude and mostly significant. The coefficient for *primetime* is always positive but not statistically significant; therefore, there is no clear evidence of an increase in the competitive mix of TV programs offered during the evening. The coefficient for *news* is always negative and significant at the 5% level, thus confirming our hypothesis that the daily news on SRF1 may attract some viewers away from a tennis match. The coefficient for *pause* is always significantly negative, indicating that some viewers rapidly switch channels during the pauses (after odd-numbered games or between the sets), possibly to skip the commercials.

To investigate the question regarding whether suspense and surprise become more important as a match progresses, we introduce two interaction variables. First, we investigate whether the effects of suspense and surprise are different among sets (*set*): playing more sets leads to slower information revelation, which might generate additional entertainment value from suspense and surprise (Ely et al., 2015). Second, we investigate whether the effect of suspense and surprise is different between the third, fourth, and fifth sets jointly ($late\_set = 1$) and the group constituted by the first two sets ($late\_set = 0$): as the earliest that matches can be won is in the third set, the first two sets might be less entertaining.[22]

Table 4 presents the results. Because our interest is in how these two interaction variables influence the effects of suspense and surprise on TV audience ratings, the table reports only the main effects and the coefficients on the interactions, i.e., any incremental impact that these factors have on the audience ratings. Also included in the specifications but not shown in the table are the control variables for the main specification.

Overall, the evidence supports our assumptions: the interaction coefficients for suspense and surprise with *set* (Panel A) are positive and significant in both models. The interaction coefficients are particularly high for surprise. For example, according to column (2), for each unit change in *set* (i.e., each additional set) the slope of the suspense on audience increases by

[22]Differently from the fourth and fifth set, the third set can end a match only when a player leads 2-0 in the third set. Thus, in a further test, we use a slightly different definition: the variable *late_set* equals one if the set number is the *3rd and* one player had a 2-0 lead, *4th*, or *5th* set, and zero otherwise. The results are similar.

Table 4
Interaction models.

Panel A: Interaction with *set*

	Dependent variable: *audience*	
	Markov	Betting odds
	(1)	(2)
SUS	5.294*	6.056*
	(2.941)	(3.562)
SUR	31.982*	11.087
	(16.832)	(9.254)
set	-2.911	-2.411
	(2.211)	(1.998)
$SUS \times set$	4.241***	6.285**
	(1.443)	(3.142)
$SUR \times set$	28.551***	16.950*
	(7.837)	(9.416)
Control variables	Yes	Yes
N	8,332	8,387

Panel B: Interaction with *late_set* (dummy)

	Dependent variable: *audience*	
	Markov	Betting odds
	(3)	(4)
SUS	29.151**	31.155*
	(11.711)	(16.914)
SUR	50.221*	32.355*
	(27.898)	(18.563)
late_set	4.476	5.653
	(3.269)	(4.352)
$SUS \times late\_set$	13.525**	22.412***
	(5.132)	(7.308)
$SUR \times late\_set$	82.920***	66.425***
	(21.734)	(23.415)
Control variables	Yes	Yes
N	8,332	8,387

Notes: The table reports the results of panel regressions with interaction terms. The dependent variable is the TV audience ratings (in thousands). The main independent variables, suspense and surprise, are derived from the Markov model (Markov) and in-play betting odds (Betting odds) and interacted with different variables. The variable *set* takes discrete values between 1 and 5. The variable *late_set* equals one if the set number is the $3rd$, $4th$, or $5th$ set, and zero otherwise. The control variables (four audience lags, primetime, news, pause, match fixed effects, and time dummies) and a constant are included but not reported. Robust standard errors that have been adjusted for clustering at the match level are provided in parentheses. In all models, *, **, and *** denote significance at the 10%, 5% and 1% levels, respectively.

approximately 6,200 viewers and the slope of surprise on audience increases by approximately 16,900 viewers.

The evidence from the second interaction model (Panel B) corroborates the notion that both suspense and surprise generate even more entertainment value during a match's later moments, i.e., when the stakes are higher. Again, the interaction coefficients are particularly high for surprise. For example, according to column (4), for each unit change in *late_set* (i.e., being in a potentially decisive set) the slope of suspense on audience increases by approximately 22,400 viewers and the slope of surprise on audience increases by approximately 66,420 viewers. Overall, the coefficient on the interactions appears to be economically non-trivial.

5.3 Robustness checks

Alternative Specifications and Postestimation Tests

In this subsection, we provide the results of further robustness checks. We begin by discussing alternative specifications and our postestimation analyses (untabulated results). First, we estimate a panel regression with one lag of audience on the right hand side of the equation. Although our model contains a lagged dependent variable, Nickell bias should not be an issue, as we work in a "*large* T, large N" context (Nickell, 1981).[23] Second, we estimate the model without the lagged audience but allowing the error term to be first-order autoregressive ($u_{i,t} = \rho \cdot u_{i,t-1} + \eta_{i,t}$, where $|\rho| < 1$ and $\eta_{i,t}$ is independent and identically distributed with a mean of 0 and a variance of σ_η^2). Finally, as the dependent variable (*audience*) is a nonnegative integer, we also use the panel Poisson regression and the panel Negative Binomial regression, both with one lag of audience on the right hand side (Winkelmann, 2013). Overall, the coefficients of suspense and surprise are always positive and highly statistically significant, with surprise being consistently larger than suspense throughout all the specifications.

Second, we control in different ways for other time effects. We estimate models including a linear and quadratic continuous time trend in the form of the elapsed time during a match and time of the day (at both the minute and hour levels). All the results are robust to these alternative specifications.

Third, we investigate whether our main results are robust to outliers in the dependent vari-

[23] "N" represents the number of cross-sectional units and "T" the number of time points.

able. We use two procedures for reducing the effects of the tails: in the first, we delete the 5% (or 1%) largest and smallest ratings; in the second, we winsorize the top and bottom 200 (or 100) observations. Regressions that use trimmed or winsorized audience figures produce similar results.

Fourth, we also check the data for multicollinearity. As discussed in Ely et al. (2015), realized suspense and surprise tend to be positively correlated, which conforms to the "intuition that more suspenseful events also generate more surprise" (Ely et al., 2015, p.245). Although multicollinearity would not reduce the reliability of the model as a whole, it might result in unstable coefficient estimates and wildly inflated standard errors. The overall sample correlation between suspense and surprise is 0.36 (0.38) for the Markov chain (betting odds) method, which is relatively low. Nonetheless, we compute the variance inflation factors (VIF) for all the dependent variables: based on the results of this analysis, we can conclude that multicollinearity is not a problem.[24]

Finally, we test the sensitivity of our results with regard to S_i, i.e., the player's probability of winning a service point. We replicate all analyses using either the historical S at Wimbledon (0.66) or the average S in our sample (0.69) for all players. The results are robust to these alternatives.

Swiss players subgroup

In an experimental study on the effects of suspense on enjoyment, Peterson and Raney (2008) note that the utility from watching a sportscast also depends on the viewer's disposition toward the participants. Disposition theory describes this effect in detail and proposes that "enjoyment derived from witnessing the success and victory of a competing party increases with positive sentiments and decreases with negative sentiments towards that party" (Zillmann, Bryant, & Sapolsky, 1989, p.162). Because our data come from Swiss households and because 35 of the matches in the sample include Roger Federer or Stan Wawrinka, both of whom are Swiss players,

[24] We check whether the VIF values are below 10, a generally accepted level indicating that multicollinearity exerts no significant impact. The VIF values for all variables in the Markov (betting odds) models range from 1.03 (1.02) to 1.55 (1.64), with a mean value of 1.26 (1.28). For running this test, we exclude the time dummies from the main regression equation specified by equation (9), as they all have a VIF of 1.00 and may deflate the mean VIF (in both models, the mean VIF would be 1.02).

the audience might show an affective disposition toward these players.[25] Notably, the mean audience for these 35 matches is three times higher than for the other 45 matches: 168,000 viewers ($\sigma = 182,140$) versus 56,733 ($\sigma = 39,823$). Although the overall audience level is not important to our analysis, it might nonetheless indicate a different minute-level TV behavior of the audience.

To address disposition theory, we therefore perform another analysis of the TV audience variation for Swiss and non-Swiss players separately. In Table 5, Panel A shows that suspense and surprise are important entertainment factors in both subsamples and that surprise has a larger effect, thus confirming our main results. However, it appears that surprise has an even larger effect when Swiss players are on the court. As Federer has regularly been very successful in the Wimbledon Championships tournament – he is tied with Pete Sampras for the most men's singles championships won (seven) – the entertainment derived from surprising moments seems to be positively amplified. Finally, the results of a regression with only a subgroup of Federer's matches are almost identical to our main results.

Audience by gender

In a study on the relationship between gender and audience experiences with televised sports, Gantz and Wenner (1991) note that men are more likely than women to become emotionally involved in sports contests and are more responsive while watching. Hence, our results might be driven solely by male viewers. Therefore, we test the validity of the results for male and female viewers separately. In our sample, the male audience (59,308 viewers) is on average higher than the female audience (47,177 viewers). In Table 5, Panel B shows the results for the subsample of female viewers (odd-numbered columns) and the results for the subsample of male viewers (even-numbered columns). The coefficients of suspense and surprise are highly significant and positive for both genders. In particular, the male audience appears to be more responsive to suspense and to surprise, confirming the idea that men are more likely to enjoy the drama and tension involved. Interestingly, columns (5) and (7) show that female viewers are also more

[25] Overall, Federer played in 30 matches and Wawrinka in six; only one match saw them play against one another. Regarding the last three stages of the tournament included in our sample, Federer played 3/6 finals, 3/12 semifinals, and 4/14 quarterfinals, whereas Wawrinka played only one quarterfinal.

markedly responsive to surprising moments than to suspenseful moments. Overall, this evidence suggests that our results are equally valid for male and female viewers.

Table 5
Robustness checks.

	Panel A: Swiss players subgroup			
	Dependent variable: *audience*			
	Markov		Betting odds	
Player group:	Swiss=1	Swiss=0	Swiss=1	Swiss=0
	(1)	(2)	(3)	(4)
SUS	21.736***	11.251***	30.997**	28.496***
	(7.967)	(2.932)	(13.410)	(10.160)
SUR	108.217***	71.021**	70.748**	47.434***
	(36.567)	(29.635)	(31.954)	(15.479)
Control variables	Yes	Yes	Yes	Yes
N	3,664	4,668	3,714	4,673
No. of matches	35	45	35	45
	Panel B: Audience by gender			
	Dependent variables: ♀ *audience*, ♂ *audience*			
	Markov		Betting odds	
Audience group:	♀ *audience*	♂ *audience*	♀ *audience*	♂ *audience*
	(5)	(6)	(7)	(8)
SUS	11.633***	21.901***	18.321***	24.765***
	(4.068)	(7.674)	(6.956)	(9.257)
SUR	38.396***	114.715***	24.067**	83.013***
	(12.224)	(22.706)	(12.007)	(21.583)
Control variables	Yes	Yes	Yes	Yes
N	8,332	8,332	8,387	8,387
No. of matches	80	80	80	80

Notes: The table reports the results of panel regressions. In Panel A we distinguish between matches with at least one Swiss player taking part in them (Swiss=1) and no Swiss player (Swiss=0). The dependent variable in Panel A is the TV audience ratings (in thousands). In Panel B, we distinguish the female audience (♀ *audience*) from the male audience (♂ *audience*) for the dependent variable. The main independent variables, suspense and surprise, are derived from the Markov model (Markov) and in-play betting odds (Betting odds). The control variables (four audience lags, primetime, news, pause, match fixed effects, and time dummies) and a constant are included but not reported. Robust standard errors that have been adjusted for clustering at the match level are given in parentheses. In all models, *, **, and *** denote significance at the 10%, 5% and 1% levels respectively.

6 Concluding remarks

6.1 Discussion

When designing entertainment content, decision makers should take into account that both suspense and surprise matter but also that surprise seems to be more important than suspense. In tennis, as suspense and surprise are exogenously determined by the rules and the players, we recognize that it would be difficult to increase either artificially. However, new rules were tested: thus, in 2015, matches without advantage scoring and with sets of first-of-four games were played (CNN, 2015). As we find that suspense and surprise are more important during a match's later moments, such rules targeted at reducing the length of matches might not be a good idea.

Furthermore, our results can be used to evaluate the format of sports competitions. Major League Soccer in the U.S., for example, is a closed league that strongly focuses on high suspense by inducing a fixed number of teams to compete with comparable levels of talent that are enforced through uniform salary caps and extensive revenue redistribution. In contrast, various European soccer leagues allow teams to be more heterogeneous in expenditures on talent and consequently in playing strength. However, relegation and promotion of European clubs based on performance merit ensures that disparities in playing strength cannot exceed certain levels within one league. Thus, the European setting increases the potential for surprise, as underdogs regularly encounter clear favorites and occasionally beat them, but does not ignore the value of suspense. Moreover, parallel European club competitions, such as the Union of European Football Associations (UEFA) Champions League, add additional suspense by matching comparably strong top European clubs.

Most importantly, our findings have implications for the design of entertainment content, particularly where suspense and surprise can be endogenously determined through a rigorous design of their contents, such as movies, TV shows and series, or talent contests. Our methodology can be applied to effectively measure how entertaining an audience perceives the product to be. The challenge remains in measuring people's beliefs. Nonetheless, even for settings in which prediction markets are unavailable or where theoretical modeling is extremely problematic, measurement of beliefs can be feasible.

For example, social media analytic tools can be used to analyze users' posts and comments on social media platforms such as Facebook. Just as tennis fans talk about and want to hear about Wimbledon on Twitter, TV series fans do the same. Alternatively, beliefs might be derived from historical data. For example, to measure the average audience's belief that a particular movie will have a happy ending, one might use as prior belief the historical fraction of movies of a certain genre that have a happy ending and compute suspense and surprise using the actual type of movie ending, i.e., the posterior belief.

Last but not least, the timing and selling of TV advertisements can also be improved. As suspense and surprise enhance entertainment and simultaneously attract human attention, commercials will be more effective following very suspenseful or surprising moments. A larger and more attentive TV audience can be reached, thereby increasing the advertisement's effectiveness. Moreover, the understanding of the effects of suspense and surprise could be translated into higher advertising revenues. For instance, broadcasters might auction advertising slots to companies willing to advertise their products. Hence, it is in the interest of content providers that bidding firms understand the impact of suspense and surprise on the audience: a slot after a very surprising moment could be sold for large amounts. Of course, as more and more entertainment content is available online, including the Wimbledon tournament at www.wimbledon.com, these implications are not restricted to TV commercials.

6.2 Conclusion

Understanding how and when enjoyment from suspense and surprise affect entertainment demand is essential for designing entertainment content. Our paper provides evidence that both suspense and surprise drive entertainment demand and become increasingly important over time. Our results also suggest that surprise matters more than suspense. We draw this inference by estimating audience beliefs and relating them to high-frequency TV audience figures in the real-world setting of the Wimbledon Championships tennis tournament.

Based on the results of this paper, we discussed important implications for sports and for the design of entertainment content in general. However, we recognize that the issue of suspense and surprise is multifaceted. Specific time patterns and combinations of suspense and surprise may produce different levels of entertainment. For example, an audience might better "tolerate" a boring moment when it follows (or precedes) a very entertaining moment; conversely, an audience might feel anxious if there are long periods with too much suspense or surprise. Because of data limitations, we were not able to precisely distinguish the effects of suspense and surprise from other factors that may also determine entertainment demand. Mood management, escapism, or learning motivations, for example, might be relevant in contexts other than sports. Ideally, detailed data at the individual level might also be used to thoroughly examine individual reactions to various levels of suspense and surprise. We leave these important subjects for future research.

Acknowledgements

The authors are grateful to the All England Lawn Tennis Club, IBM, and the *Schweizer Radio und Fernsehen* for providing the data. We are indebted to Alex Frankel for useful consultation. The authors thank seminar participants at the University of Zurich and conference participants at the 7*th* European Sport Economics Association (ESEA) Conference in Zurich for very helpful comments and suggestions. Finally, we thank Fabio Banfi for helping with Matlab and Python. We take responsibility for all remaining errors.

Appendix A

This appendix introduces the basic rules of tennis (Part I) and explains how we compute the probability of winning a point on service (Part II).

Part I Player 1 begins the match by serving in the first game of the first set. Player 1 wins a *point* (sometimes referred as "point game") if player 2 cannot return the ball. A *game* is won when one player wins four points with a two-point difference, or when there is a two-point difference after a deuce, i.e., a score of 40–40 (3 points to 3 points in a game). The players alternate serving every game, and they change ends after every odd-numbered game. A *set* is won when a player either wins six games with a two game difference, or, in the case of a tie-break when the score for one player is 7:6. The *tie-break* begins when the game score is tied at 6:6, and is played until one player wins seven points with a two-point difference, or until there is a two-point difference when the point score is 6-6. At Wimbledon, a tennis match is played as the best of five sets (instead of three), meaning that a player winning three sets wins the match. The fifth set does not have a tie-break; the set is won when one player has won two games more than the other. For further details on the rules of tennis, please consult: www.itftennis.com/media/107013/107013.pdf.

Part II To compute the probability of winning a point on service for each player in our sample using historical data, we download the necessary statistics from www.atpworldtour.com, where official player-level data are available. For all the 160 server-match combinations in our sample (80 matches, two players per match), we download the player's average "*% 1st service in*", "*% of points won if 1st service in*", and "*% of points won if 2nd service in*" statistics from the previous year on grass courts. For players with no history on grass courts for the previous year, we use statistics from two years earlier or, when also unavailable, from the previous year but on hard courts. Then, we apply the formula provided by Klaassen and Magnus (2014, p.75):

$$
\begin{aligned}
S_i = &(\textit{\% 1}^{st}\textit{ services in}) \cdot (\textit{\% of points won if 1}^{st}\textit{ service in}) \\
&+ (\textit{\% 1}^{st}\textit{ services not in}) \cdot (\textit{\% of points won if 2}^{nd}\textit{ service in})
\end{aligned}
\tag{10}
$$

The first part of the formula reflects the probability of winning the point on the first serve, whereas the second part reflects the probability of winning the point on the second serve when the first serve is faulted. Thus, S_i reflects the fact that a player can win a point on either the first *or* second service. As an illustration, we report a numerical example provided in Magnus and Klaassen (1999) using Wimbledon data on men's singles from 1992 to 1995: $S_i = 0.587 * 0.777 + (1 - 0.587) * 0.518 = 0.456 + 0.214 = 0.67$. In our sample, $S_i = 0.69$ on average: Federer has the highest S_i (0.787 in 2009), whereas Albert Ramos-Vinolas has the lowest (0.483 in 2012). Overall, the advantage of this procedure over using a fixed S for all matches is that we use the same statistics that are also readily available to gamblers.

Appendix B

This appendix provides details on the Mediapulse TV panel.

- The TV audience measurement system: every household in the panel is given a small measuring device that is connected to all TV sets in the house or apartment. This field-tested device is used in almost 25 countries worldwide. It collects audience information every second. Mediapulse then aggregates these data for the entire panel and saves it at the minute-by-minute level. For example, the TV audience corresponding to the minute 15:23:00 indicates the average audience between 15:23:00 and 15:23:59.

- Extrapolation: the reporting samples of a panel are extrapolated to the population (universe) estimates, which allows the results to be representative of the respective population, i.e., of the TV panel target audience (Kuonen & Hulliger, 2013). For extrapolating, Mediapulse uses actual population figures from the Swiss Federal Statistical Office. The quota attributes used to determine the appearance of a household in the panel are language area, canton (a member state of Switzerland), district (a region of a canton), household size, presence of children in the household, and age of the head of household. To certify that the panel conforms to international quality standards, it is subject to an annual external verification.

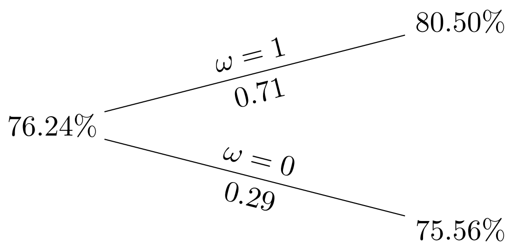
- Changes in the panel: over the 2009-2014 period, two changes in the measuring system have occurred. First, beginning in 2010, daily weighting was introduced, and the use of Teletext was considered to be TV watching. Second, before 2013, a panel of 1,918 households was randomly recruited by the phone. This universe comprised all households with at least an installed TV. However, after January 1, 2013, the panel was recruited by either the phone or mail. It now contains 2,000 households, at least 1,870 of which must provide data daily. For both panels, households watching TV exclusively from a computer are not included. Mediapulse informed us that for an audience analysis *within* a program, like a tennis match, none of these changes has had an impact on the size of the variation in the measured audiences.

Appendix C

This appendix illustrates how to compute SUS^{Markov} and SUR^{Markov} (Part I) and provides further detailed descriptive statistics for the main variables in our model (Part II).

Part I Using a fictitious example, we illustrate how to compute SUS^{Markov} and SUR^{Markov}, i.e., suspense and surprise based on a finite binary Markov chain. An almost identical procedure applies to computing SUS^{odds} and SUR^{odds}. Player W is playing against player L. After the $63rd$ point, the match is tied at one set apiece, with player W ahead five games to four (5-4) in the set, and 40 points to zero (40-0) in the game. Therefore, player W is serving to win the game and, in so doing, also the set (the expression used is: "Player W has three set points"). From his historical serving records on grass, we compute server W's probability of winning the next point, $S_W = 0.71$ (see Part II of Appendix A). As the situation of the players in a tennis match is always symmetrical, player L's probability of winning on a return point is 0.29.

Using this information, the model predicts that player W's match winning probability is 76.24% (i.e., the current belief μ_p). If he wins the point ($\omega = 1$), the posterior belief would rise to 80.50% (+4.26%). However, as there is a 29% chance that he will lose next point ($\omega = 0$), that loss would bring the posterior belief down to 75.56% (-0.68%). The probability transition from point 63 to point 64 is:



where 80.50% and 75.56% are the posterior beliefs. The possible size of the update in the beliefs is correctly asymmetric (+4.26% vs. -0.68%), as player W would still have two set points left to play even if he loses this serve point.

Finally, if player W indeed loses the $64th$ point ($\omega = 0$), the posterior belief would drop to 75.56%, making 80.50% the counterfactual probability. Using equation (2), we compute suspense for the 63rd point ($p = 63$) as the standard deviation of the next point's beliefs:

$$SUS_{63}^{Markov} = [0.71 \cdot (0.805 - 0.7624)^2 + 0.29 \cdot (0.7556 - 0.7624)^2]^{1/2} = 0.036.$$

Similarly, using equation (3) we compute surprise for the 64th point ($p = 64$) as the absolute value of change in beliefs from point 63 to 64:

$$SUR_{64}^{Markov} = |0.7556 - 0.7624| = 0.0068.$$

Thus, the model predicts a suspense of 0.036 and a surprise of 0.0068.

Part II

Table A1
Descriptive statistics of suspense and surprise by tournament stage.

			Markov		Betting odds	
Tournament stage	N	*audience*	*SUS*	*SUR*	*SUS*	*SUR*
1*st* stage	12	44.0	0.0064	0.0061	0.0131	0.0093
		(29.23)	(0.0121)	(0.0110)	(0.0231)	(0.0175)
2*nd* stage	14	55.2	0.0096	0.0078	0.0173	0.0098
		(44.35)	(0.0208)	(0.0118)	(0.0362)	(0.0179)
3*rd* stage	12	75.8	0.0103	0.0084	0.0193	0.0125
		(45.16)	(0.0407)	(0.0175)	(0.0431)	(0.0193)
4*th* stage	10	83.2	0.01284	0.0096	0.0216	0.0127
		(84.85)	(0.0413)	(0.0177)	(0.0432)	(0.0198)
Quarterfinal	14	87.4	0.0193	0.0154	0.0281	0.0138
		(59.84)	(0.0477)	(0.0195)	(0.0475)	(0.0204)
Semifinal	12	89.1	0.0209	0.0162	0.0320	0.0162
		(54.83)	(0.0502)	(0.0205)	(0.0506)	(0.0205)
Final	6	362.8	0.0224	0.0199	0.0348	0.0167
		(252.38)	(0.0527)	(0.0206)	(0.0582)	(0.0218)

Notes: Displayed are summary statistics for the TV audience ratings (in thousand), suspense, and surprise for the 80 men's singles matches played at Wimbledon between 2009 and 2014, for a total of 8,563 minutes of live tennis. Standard deviations are provided in parentheses.

References

Alavy, K., Gaskell, A., Leach, S., & Szymanski, S. (2010). On the edge of your seat: Demand for football on television and the uncertainty of outcome hypothesis. *International Journal of Sport Finance*, *5*(2), 75-95.

Alwitt, L. F. (2002). Suspense and advertising responses. *Journal of Consumer Psychology*, *12*(1), 35–49.

Arellano, M., & Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of econometrics*, *68*(1), 29–51.

Baldi, P., & Itti, L. (2010). Of bits and wows: A bayesian theory of surprise with applications to attention. *Neural Networks*, *23*(5), 649–666.

Barnett, T., & Clarke, S. R. (2005). Combining player statistics to predict outcomes of tennis matches. *IMA Journal of Management Mathematics*, *16*(2), 113–120.

Borland, J., & MacDonald, R. (2003). Demand for sport. *Oxford Review of Economic Policy*, *19*(4), 478–502.

Bryant, J., Rockwell, S. C., & Owens, J. W. (1994). Buzzer beaters and barn burners: The effects on enjoyment of watching the game go "down to the wire". *Journal of Sport & Social Issues*, *18*(4), 326–339.

CNN. (2015). *Fast4: Tennis' answer to cricket's twenty20.* Retrieved from http://edition.cnn.com/2015/01/12/tennis/fast4-new-tennis-format/ (Retrieved on 13 May, 2016)

Comisky, P., & Bryant, J. (1982). Factors involved in generating suspense. *Human Communication Research*, *9*(1), 49–58.

Croxson, K., & Reade, J. (2014). Information and efficiency: Goal arrival in soccer betting. *The Economic Journal*, *124*(575), 62–91.

Ely, J., Frankel, A., & Kamenica, E. (2015). Suspense and surprise. *Journal of Political Economy*, *123*(1), 215–260.

Forrest, D., Simmons, R., & Buraimo, B. (2005). Outcome uncertainty and the couch potato audience. *Scottish Journal of Political Economy*, *52*(4), 641–661.

Gantz, W., & Wenner, L. A. (1991). Men, women, and sports: Audience experiences and effects. *Journal of Broadcasting & Electronic Media*, *35*(2), 233–243.

Hasbrouck, J. (1991). Measuring the information content of stock trades. *The Journal of Finance*, *46*(1), 179–207.

Horstmann, G. (2015). The surprise–attention link: A review. *Annals of the New York Academy of Sciences*, *1339*(1), 106–115.

Hutchins, B. (2014). *Game, set, cash! inside the secret world of international tennis trading.* Black Inc.

Itti, L., & Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision research*, *49*(10), 1295–1306.

Klaassen, F., & Magnus, J. R. (2014). *Analyzing Wimbledon: The power of statistics.* Oxford University Press, USA.

Kuonen, D., & Hulliger, B. (2013). *Evaluation of the new mediapulse television panel with respect to its suitability for local television channels* (Tech. Rep.). Mediapulse AG.

Liu, Y. (2001). Random walks in tennis. *Missouri Journal of Mathematical Sciences*, *13*(3).

Magnus, J. R., & Klaassen, F. J. (1999). On the advantage of serving first in a tennis set: four years at Wimbledon. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *48*(2), 247–256.

Newton, P. K., & Aslam, K. (2009). Monte carlo tennis: A stochastic markov chain model. *Journal of Quantitative Analysis in Sports*, *5*(3).

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica: Journal of the Econometric Society*, 1417–1426.

Olson, J., & Stone, D. F. (2014). Suspense-optimal college football play-offs. *Journal of Sports Economics*, *15*(5), 519–540.

O'Malley, A. J. (2008). Probability formulas and statistical analysis in tennis. *Journal of Quantitative Analysis in Sports*, *4*(2).

Peterson, E. M., & Raney, A. A. (2008). Reconceptualizing and reexamining suspense as a predictor of mediated sports enjoyment. *Journal of Broadcasting & Electronic Media*, *52*(4), 544–562.

Rodríguez, C., Pérez, L., Puente, V., & Rodríguez, P. (2015). The determinants of television audience for professional cycling the case of Spain. *Journal of Sports Economics*, *16*(1), 26–58.

Su-lin, G., Tuggle, C. A., Mitrook, M. A., Coussement, S. H., & Zillmann, D. (1997). The thrill of a close game: Who enjoys it and who doesn't? *Journal of Sport & Social Issues*, *21*(1), 53–64.

Vorderer, P., Klimmt, C., & Ritterfeld, U. (2004). Enjoyment: At the heart of media entertainment. *Communication theory*, *14*(4), 388–408.

Vorderer, P., Wulff, H. J., & Friedrichsen, M. (2013). *Suspense: Conceptualizations, theoretical analyses, and empirical explorations.* Routledge.

Walker, M., Wooders, J., & Amir, R. (2011). Equilibrium play in matches: Binary markov games. *Games and Economic Behavior*, *71*(2), 487–502.

Winkelmann, R. (2013). *Econometric analysis of count data.* Springer Science & Business Media.

Wolfers, J., & Zitzewitz, E. (2006). *Interpreting prediction market prices as probabilities.* (Working paper)

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data.* MIT press.

Zillmann, D. (1991). The logic of suspense and mystery. *Responding to the screen: Reception and reaction processes*, *7*, 281–303.

Zillmann, D. (1996). The psychology of suspense in dramatic exposition. *Suspense: Conceptualizations, theoretical analyses, and empirical explorations*, 199–231.

Zillmann, D., Bryant, J., & Sapolsky, B. S. (1989). Enjoyment from sports spectatorship. *Sports, games, and play: Social and psychological viewpoints*, *2*, 241–278.