Passive vs. Active Attention to Baseball Telecasts: Implications for Content (Re-)Design*

Xiao Liu† Matthew Shum‡ Kosuke Uetake§

Abstract

Do TV program features affect consumer viewing and attention behaviors differently? How should a planner design TV contents to increase viewer engagement to programs and commercials? Using unique individual-level data containing high-frequency logs which detail whether viewers are *passively* or *actively* paying attention while watching TV, we study how game-play features, including suspense and surprise, influence viewers' attention levels. Overall, only a small fraction of viewers are actively paying attention, and viewers value suspense over surprise. Viewers pay less attention during commercials, but they do not walk away from the TV or "zap" to another station. These results have implications for content design, as reshuffling commercials to the most suspenseful moments in the game attracts more attention. While shortening baseball games has ambiguous effects on attention, a "mercy rule" (which selectively shortens less competitive games) has sizeable effects on attention.

*First draft: June 18, 2020. This version: March 23, 2021. We thank Eric Anderson, Colin Camerer and his lab team, Avi Goldfarb, J-P Dube, Brett Hollenbeck, Daniel Ershov, Yufeng Huang, Ed Kaplan, Scott Kaplan, Tom Meyvis, Ilya Morozov, Leif Nelson, Aniko Oery, Andrey Simonov, K. Sudhir, Anna Tuchman, Raluca Ursu, Caio Waisman, Larry White, and Ken Wilbur for helpful comments. We also thank the seminar and conference participants at Marketing Science Conference, Northwestern Kellogg, Virtual Digital Economy Seminar, Yale SOM. The paper's results are our owns and do not represent the companies' views.

†New York University. `xliu@stern.nyu.edu`

‡California Institute of Technology. `mshum@caltech.edu`

§Yale School of Management. `kosuke.uetake@yale.edu`

1 Introduction

Entertainment plays an important role in our life. An average American adult spends about 3 hours watching TV every day (American Time Use Survey 2018). People watch news for obtaining useful information such as weather and stock prices, and watch drama and sports for enjoying leisure. During the ongoing COVID-19 crisis, people are spending even more time watching television programming. A recent study by Nielsen says streaming of TV content continues to post nearly double the levels of a year ago.[1] At the same time, while television viewing has increased, industry analysts are worried that attention levels while viewing may have dropped due to viewers' ready access to smartphones or tablets, which distract from the television programming. Hence, it is challenging to measure and understand audience engagement in the current era when so many alternative distractions are present.

In this paper, we study the relationship between TV audience engagement and the contents of television programming and consider how to (re-)design program contents to optimize audience engagement. We utilize a unique dataset containing high frequency (second-by-second) logs of not only whether individuals are tuned in to a program, but also their level of attention – hence, for instance, we can distinguish whether viewers are *actively* attentive (their eyes are focused on the TV screen) or *passively* attentive – sitting in front of the TV but not watching the program.

Using this data, we estimate a choice model where agents choose both *whether* or not to watch TV and *how much* attention to pay, depending on the current contents of the broadcast. Our empirical context is baseball telecasts during the 2018 Japanese major league season. Baseball games are an ideal laboratory for our research for multiple reasons. First, professional baseball games are very popular in Japan as in the U.S. Hence, TV channels regularly broadcast baseball games and a large number of viewers

[1] https://www.mediapost.com/publications/article/351299/streaming-tv-usage-still-strong-vs-2019-but-sli.html

watch the games on TV.[2] Second, game content and status can be easily summarized by a low dimensional set of state variables, such as innings, outs, bases loaded, score differences, etc. This feature allows us to summarize baseball content relatively easily and credibly compared to, for example, football or drama. Third, heterogeneity across games is substantial. Some games are boring as one team scores 10 runs in the first inning and dominates the other team, whereas some games are exciting as the game ties until the last inning. The rich heterogeneity across games helps us identify viewers' preferences over game contents.

We find that features that describe the excitement of gameplay, including *suspense* and *surprise* (Ely, Frankel, and Kamenica 2015), affect decisions to pay passive and active attention differently. Overall, suspense has positive and significant effects on viewing and attention, but not surprise. This corroborates findings from the previous literature. Moreover, suspense also modulates viewers' attention levels during commercials which air during the telecasts.

Using these results, we simulate several counterfactual scenarios to assess the revenue impact of rule changes in baseball, as the largest component of baseball revenue derives from television. First, we find that reshuffling commercials to the most suspenseful moments increases attention. Additionally, we find that shortening baseball games has ambiguous effects on attention, but a "mercy rule" which essentially, selectively shortens less competitive games, leads to sizeable increases in attention, and (surprisingly) more so to commercials than to the game. Thus, overall, we find that these proposed changes have important implications for baseball teams and for advertisers. The latter two counterfactual scenarios resemble actual proposed rule changes which have been discussed by Major League Baseball in the United States, especially for the Covid-19 shortened 2020 season.[3]

Although prior research has also looked into the relationship between program

[2] https://www.nielsen.com/us/en/insights/article/2020/during-covid-19-sports-viewers-are-still-a-scoring-opportunity-for-brands-and-media-owners/

[3] https://www.cbssports.com/mlb/news/10-rule-changes-mlb-could-test-during-the-shortened-2020-season-including-universal-dh-and-a-mercy-rule/

contents and audience engagement, several areas remain unexplored. First, the dominant engagement metric used in prior studies is *viewing*, measured traditionally by set-top boxes installed in homes of a representative sample of households.[4] This may be a rather crude measure because consumers often multi-task and may not pay full attention to the television programming. This problem may be particularly exacerbated for TV commercial breaks, during which viewers are likely to look away from the TV to chat with friends, check emails, and so on. Second, the extant research by and large only investigates engagement to programs, but ignores TV commercials (or overlooks the interactions between programs and commercials), which play an important role in the business model of TV stations.

Our paper fills these research gaps. First of all, the availability of granular data on attention in addition to viewing distinguishes our study from previous studies. The rise of the Internet and universal broadband in most industrialized countries has enabled the collection of more sophisticated measures of viewers' active engagement. The attention data we use here is collected via set-top sensors which detect individuals' faces and eye movements, allow us to distinguish whether viewers are passively or actively watching a program. As we will see, while passive and active attention inevitably move together, they differ substantially and respond differently to program features and to commercials.

Second, our high-frequency data also permit us to study the spillovers between viewer engagement towards the television programming and the commercials that air during the program. We find evidence of rich spillover effects, as increased attention to a television program may be either diluted or amplified during concurrent commercial breaks depending on the counterfactual we consider. Our focus on these spillovers between programs and commercials appears novel in the literature and provides important managerial implications for TV stations and advertisers.

[4]For example, see, http://en-us.nielsen.com/sitelets/cox/documents/NielsenandCox.pdf?lang=en

Related Literatures

Measuring Attention: visual fixations. Using AC Nielsen data (or similar set-top-box data) that measure whether people “tune” into a particular program, i.e., "TV Rating", many studies have examined consumer program choices and switching decisions (Wilbur, 2008; Goettler and Shachar, 2001; Danaher, 1995; Deng and Mela, 2018), as well as ad avoidance behaviors (Schweidel and Kent, 2010; Teixeira, Wedel, and Pieters, 2010; Yao, Wenbo, and Chen, 2017). Collectively, they identified factors that affect TV viewing behaviors, including viewer-related factors such as demographics, program-related factors such as program genre, cast demographics and commercial length, as well as time-related factors, such as commercial location.

However, this literature suffers from two limitations. First, the rating metric can be a poor measurement of consumer attention. For example, Gunter, Furnham, and Lineton (1995) found that when the TV is turned on, family members are absent for substantial proportions of the time, and even when present they do not pay active attention to the screen for more than a fraction of that time. These behaviors occurred most often during programming that attracted less visual attention, particularly ads. Second, they focus on viewing choices across programs, instead of moment-to-moment choices within a program.

We address these two limitations in this paper. For the first limitation, we leverage novel data that directly measure consumer attention using eye fixation and facial expressions.[5] The same dataset has been used in only one other marketing paper, McGranaghan, Liaukonyte, and Wilbur (2019), which studies a different research question about how contents in TV commercials affect viewability and attention. For the second one, we construct factors that represent content dynamics within a program, and quantify their impact on consumer attention choices minute-by-minute.

[5] See Hutchinson, Lu, and Weingarten (2017) for a study where eye-tracking data is used to understand consumer attention to advertisements and shopping assortments. Eye-tracking has also been used in economic studies; for instance, Wang, Spezio, and Camerer, 2010; Knoepfle, Wang, and Camerer, 2009 use eye-tracking for testing theories of learning and deception in games.

Summarizing Program Content: Suspense and Surprise. Among many factors that affect consumer preference on entertainment, two key determinants, suspense and surprise, have been recently formalized by both theoretical and empirical literature. In a seminal work, Ely, Frankel, and Kamenica (2015) propose a theoretical model of Bayesian persuasion where the audience derives entertainment utility from suspense and surprise; they define suspense as the uncertainty of future belief relative to the current belief, and surprise as the difference between the current and the prior belief, which we closely follow.

The theory is motivated by prior laboratory experiment findings associated with suspense (Bryant, Rockwell, and Owens, 1994; Su-lin, Tuggle, Mitrook, Coussement, and Zillmann, 1997; Peterson and Raney, 2008) and surprise (Itti and Baldi, 2009; Alwitt, 2002). The concept of suspense is related to the "outcome uncertainty hypothesis" in the sports economics literature (Rottenberg, 1956; Borland and MacDonald, 2003).

Empirically, several papers have investigated the relationship between suspense and surprise and TV rating in the sports context.[6] Specifically, Bizzozero, Flepp, and Franck (2016) study Wimbledon tennis matches, Buraimo, Forrest, McHale, and Tena (2020) study English Premier League soccer, and Kaplan (2020) looks into professional basketball in the United States. A common theme in this strand of research is that they used aggregate TV rating data to measure consumer engagement. Interestingly, they documented contradictory results on the relative importance of suspense and surprise, where Bizzozero, Flepp, and Franck (2016) showed that surprise has a larger impact than suspense, but Kaplan (2020) and Buraimo, Forrest, McHale, and Tena (2020) found that surprise induces a smaller viewership response. Unlike these studies, we use individual-level attention data to measure consumer response and examine a different context, baseball games. Our focus on game design strategies in the counterfactual analysis is also new.

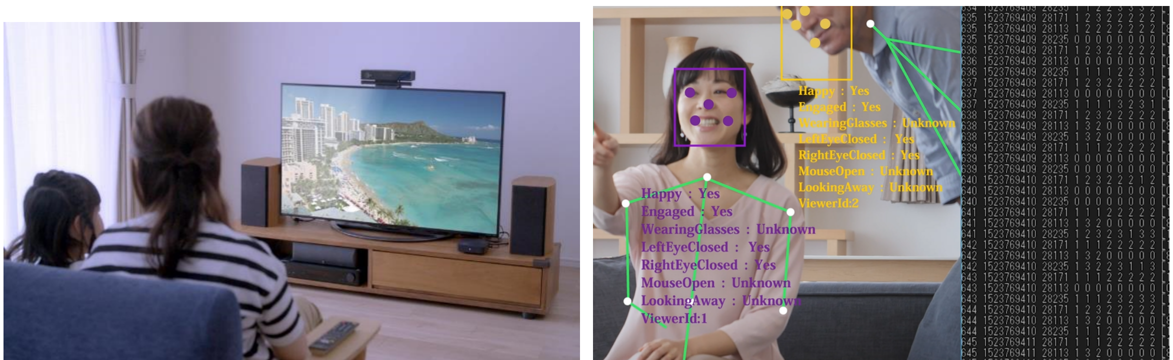
[6] An exception is a recent paper by Simonov, Ursu, and Zheng (2021), which looks at the impacts of suspense and surprise in e-sports viewing on twitch.tv. They also find that suspense is more important than surprise for viewing.

2 Data

Here we summarize the different data sources used in this paper.

2.1 Attention Data

Figure 1: Device and Facial Recognition System



(a) (b)

Note: The images are provided by TVISION INSIGHTS.

We use individual-level TV viewership and attention data from TVISION INSIGHTS (hereafter TVISION). The company creates the TVISION panel, similar to the Nielsen TV panel, by recruiting approximately 1,900 individuals in 800 households in Japan. Once a household agrees to join the panel, TVISION installs a device with a facial recognition capability to track each household member's TV viewing behavior.

The device gathers rich second-by-second data on the television engagement behaviors of household members at three levels. First, the device detects the TV programs the household tunes in and the advertisements during each program. Second, when the TV is on, the device detects whether a household member is in front of the TV or not. Thanks to the facial recognition algorithm, it can differentiate which household member is in front of the TV. Finally, among the household members situated in front of the TV set, the device collects whether each member is actively paying attention to the TV screen. Again, the facial recognition algorithm allows us to tell whether each

Table 1: Summary Statistics: Individuals

	During each minute, percentage of viewers who:		
	Not tuned in to baseball	Passive attention	Active attention
During Commercial Minute	92.91%	5.84%	1.24%
During Gameplay Minute	93.76%	4.96%	1.28%
During all minutes	93.63%	5.1%	1.27%

member pays attention to the TV screen or not.

The left picture in Figure 1 shows how the device is installed in each household. A key difference between the current industry practice of collecting viewership data and TVISION's way is that people do not have to keep pushing the button to indicate they are watching TV, and hence the information is more precisely and passively collected. The right figure is an example of how the Deep Learning based facial recognition system identifies each individual in the household separately, and measures whether she/he is in front of the TV or actively pays attention to it.

Based on this rich high-frequency data, we construct *three mutually exclusive states* of an individual's television engagement each second: (i) she may not be watching TV; (ii) she is sitting in front of the TV but engaged in another activity; and (iii) she is actively watching TV. The typical "viewing" measure used by, for example, Nielsen, combines both states (ii) and (iii), and does not distinguish between passive vs. active watching. As we will see, this is an important distinction, as typically fewer than half of viewers are actively attentive at any point in time, and this fractions varies substantially depending on program features and whether a commercial is airing.

For each game in our viewing sample, we include panelists who watched at least one second of the game. Hence, we drop all panelists who did not watch any baseball telecasts in 2018. This leaves us with 1151 panelists in our sample: 52% are male, and the average age is 39.6.

In Table 1, we report summary statistics on the viewing behavior for our sample. The granularity of the original data is at the viewer-second level, and we aggregate the data to the viewer-minute level for each game, and construct minute-by-minute variables denoting what fraction of each minute of the baseball telecast a household spends in front of the TV and, additionally, what fraction of each minute a person actively pays attention. Based on the minute-level data, we further define the outcome variable which describes agents' TV viewing behavior. For each minute, we classify a viewer as "not tuned in to baseball" is she spends most of that minute either away from the TV or not tuned into the baseball game. Similarly, we classify her as "passively attentive" if she spends most of the minute in front of the TV but her eyes are not on the screen, and "actively attentive" if she spends most of the minute with her eyes focused on the TV screen.

Interestingly, passive vs. active attention levels behave asymmetrically, and do not always move together. A larger fraction of viewers are passively attentive to commercials than to the game (5.84% vs. 4.96%) but a slightly *smaller* fraction of viewers are actively attentive to commercials than to the game (1.24% vs. 1.28%). This finding that active and passive attention do not always move together implies that viewers' attention choices cannot be captured by "single index" discrete-choice models (such as the ordered logit or probit models); rather, for our empirical work, we utilize the multinomial logit model, in which separate indices determine agents' choices to pay passive versus active attention.

2.2 Baseball Data

In Japan, baseball is the most popular professional sport. In 2019, there were an average of 30,929 spectators per game (for comparison, an average US Major League Baseball game had only 28,317 spectators), and baseball games are frequently aired on TV.

We obtained detailed pitch-level data on Japanese professional baseball games (NBL) from Data Stadium Inc. The data contains the time stamp of each pitch (or steal) and its results, such as a swing-out and a home-run. We use the data of all the games in 2018, and in particular, all games that are broadcast on major TV networks during 2018.

Table 2: Summary Statistics: Baseball Games

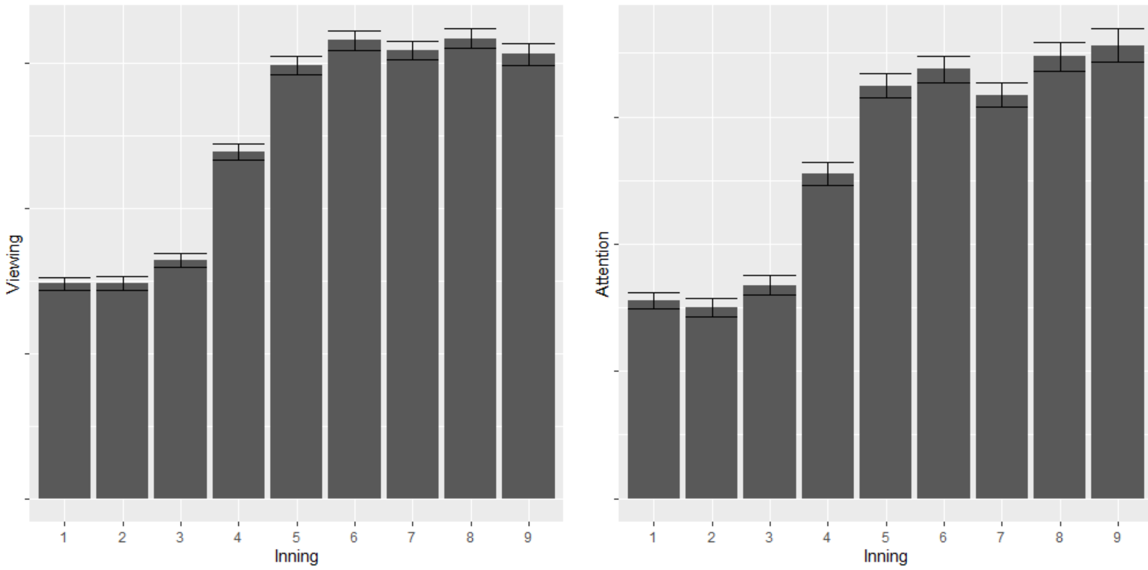
	Mean	Sd	Min	Max	N
All					
Home team win	.424	.494	0	1	877
Home team score	4.179	3.039	0	20	877
Away team score	4.365	3.087	0	20	877
Number of pitches	304.4	38.32	167	475	877
Game length (second)	12002.1	3062.8	7623	86371	877
Suspense	.035	.055	0	.709	67411
Surprise	.017	.050	0	.700	66534
Broadcasted only					
Home team win	.561	.502	0	1	41
Home team score	4.415	2.966	0	11	41
Away team score	3.293	2.562	0	11	41
Number of pitches	297.1	35.3	232	386	41
Game length (seconds)	11544.3	1746.6	8964	16646	41

Note: The unit of analysis for Suspense and Surprise is at the minute level.

Table 2 reports summary statistics of all the games in 2018. Among 877 games in total, the home team wins a game with 42.4% probability, and scores about 4.2 points on average. There are about 300 pitches in each game and each game lasts on average 12,000 seconds. Similarly, the bottom part of Table 2 reports summary statistics of all the games that are broadcast in 2018. Among 41 games, the home team wins with 56.1% probability and scores on average 4.4 points. There are about 300 pitches and each game lasts on average 11,500 seconds. Hence, the home team is more likely to win when the game is broadcast, while other characteristics are similar among games.

To see when people view and pay attention to the baseball games, Figure 2 shows that people are more likely to view and pay attention during later innings, when the outcome of the game is closer to being resolved. We also find that both passive and active attention increase when the score difference between the teams is smaller, indicating that viewers are more engaged when the game is more competitive.

Figure 2: Viewing and Attention by Inning



(a) (b)

Note: The figure plots the average viewing (sitting in front of the TV but not paying attention) and attention (paying active attention) over innings using the second-by-second data. Due to an NDA, we are not allowed to show the scale of the vertical axis. Hence, the scale of the two figures are not the same.

3 Defining Gameplay Variables

Using the baseball game data, we construct a number of variables to summarize what is happening in the game at a minute-by-minute level. These include *difrun* (the run difference between the home and away teams), as well as dummy variables for the inning, and whether it is a postseason game. Following the recent literature, (notably Ely, Frankel, and Kamenica (2015)) we construct two measures of non-instrumental information that may affect people's TV viewing behavior: suspense and surprise.

We assume that as they watch the game, viewers form and update beliefs regarding which team will win the game. Specifically, at "at-bat" t in the baseball game, we let μ_t denote the perceived probability that the home team will win the game, which can depend on state variables S_t of the game.

Following Ely, Frankel, and Kamenica (2015) and Bizzozero, Flepp, and Franck (2016), we define the suspense measure as

$$SUS_t = E_t\left[(\mu_{t+1} - \mu_t)^2\right]^{1/2}. \tag{3.1}$$

The expectation is calculated with respect to the state transition, $P(S_{t+1}|S_t)$. Suspense is an "ex-ante" measure which is increasing in the variance of the change in the home-win probability between the next and current periods. In contrast, surprise is an "ex-post" quantity which arises from unexpected or unanticipated events, defined as the absolute distance between the current and previous beliefs:[7]

$$SURP_t = |\mu_t - \mu_{t-1}|. \tag{3.2}$$

In order to compute the suspense and surprise measures using the baseball data, we first need to define the state variables for the baseball game, and estimate the empirical transition probability for these state variables from the data. Complete details on this procedure are contained in the Appendix.

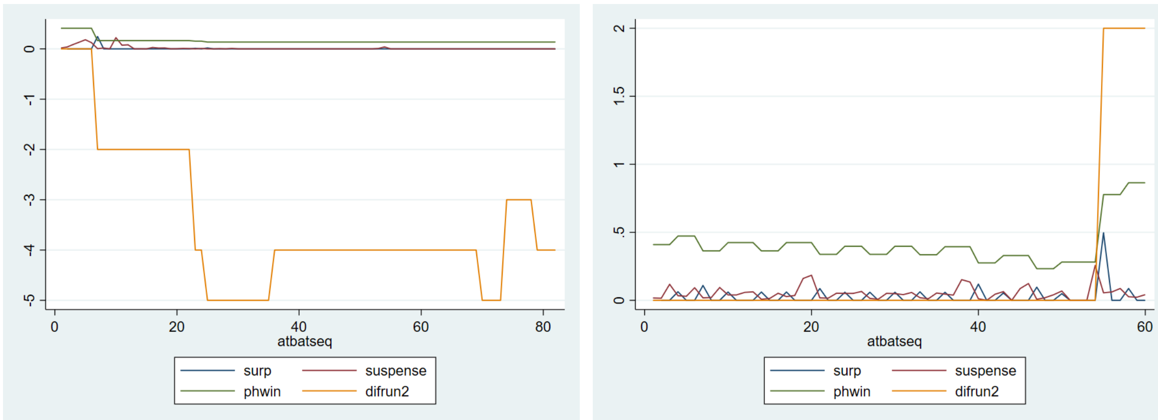
Summary Statistics of Suspense and Surprise Table 2 reports summary statistics of the suspense and surprise measures. Note that we use the data of all games in 2018 regardless of whether the game is on air. The average suspense is 0.035, while the average surprise is 0.017. Hence, suspense is twice as big as surprise on average. By contrast, the variance of the two variables look quite similar.[8]

Two subfigures in Figure 3 show how the two measures (suspense and surprise) move within a game for two exemplar games. In the first game (panel (a)), the away team scores two points in the first half of the game, then adds more points later in the

[7]We follow the specification used in Bizzozero, Flepp, and Franck 2016. Our empirical results are unchanged if we measure surprise using the Euclidean norm ($((\mu_t - \mu_{t-1})^2)^{1/2}$) instead of the absolute value.

[8]The number of observations for the surprise measure is smaller than the one for the suspense measure because the surprise measure is calculated ex-post, and hence it is not defined for the first at-bat of the game.

Figure 3: Examples of Suspense and Surprise Over Time



(a) (b)

Note: phwin is the probability of home team winning, difrun2 is the run difference between the home and away teams.

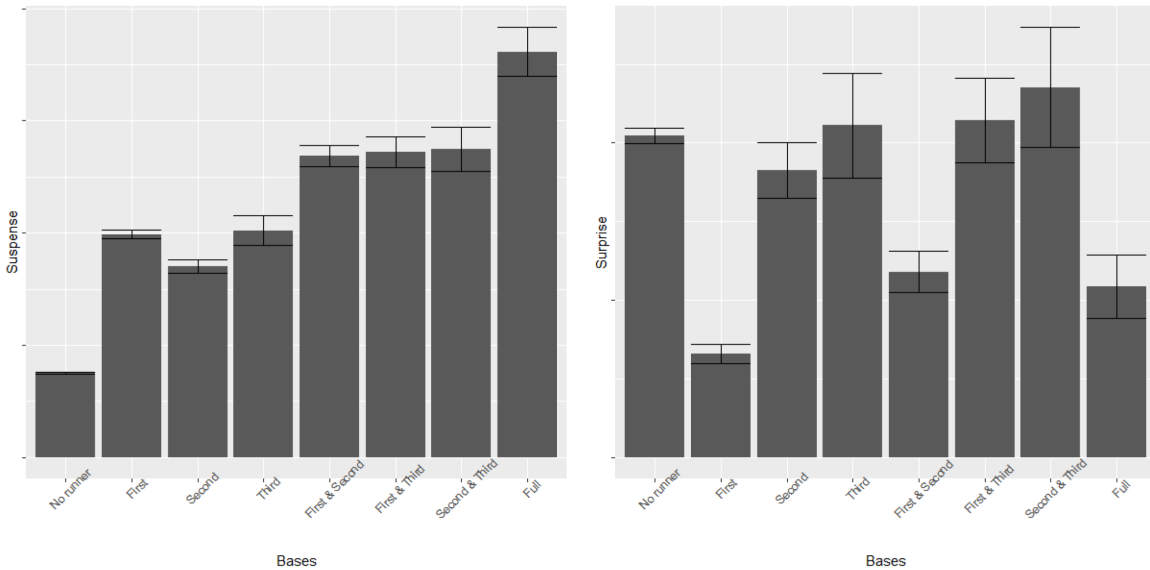
game. Hence the probability that the home team wins the game remains very small during the entire game, leading to very little suspense and surprise.

By contrast, the game shown in panel (b) of Figure 3 is a more exciting one. The game is a tie until the very end of the game, and the home team gains two points in the last inning. Although the game is scoreless until the last inning, both teams have a chance to win the game. Thus, the probability that the home team wins moves up and down, which leads to large variations in the two measures.

Figure 4 highlights some key differences in what is measured by suspense vs. surprise. The left-hand side panel shows that suspense is maximal at game junctures when the bases are loaded; this shows how suspense is an "ex-ante" measure which is highest during moments when there is a lot at stake. In contrast, we see, from the right-hand side panel, that surprise, which is an "ex-post" measure, can be very high even when no batters are on base, and is relatively low when the bases are fully loaded.

At the same time, we find that suspense and surprise move together with other game features. For example, the absolute difference in score between the competing teams is negatively correlated with both suspense and surprise, implying that both suspense and surprise are greater when the game is a tie.

Figure 4: Suspense and Surprise by Whether Batters are On-Base



(a) (b)

Note: The figures plot the average and the 95% confidence intervals for suspense and surprise by whether batters are on-base

4 Viewing and Attention Choice Model

For modeling agents' viewing and attention choices during the baseball game, we consider a multinomial logit model. While our attention data varies across each second of the telecast, we aggregate it up to the minute level to match the gameplay variables, which typically vary minute-by-minute.

In each minute t of a baseball game, an agent i takes one of the following three choices:

$$y_{it} = \begin{cases} 0 & \text{if not tuned in to baseball telecast} \\ 1 & \text{if passively paying attention} \\ 2 & \text{if actively paying attention.} \end{cases} \tag{4.1}$$

The choice y_{it} describes the most frequent action that the agent takes where $y_{it} = 0$ means that the agent is not tuned into the baseball game (either she is watching another program or the TV is off); $y_{it} = 1$ if the agent spends most of minute t (> 30 seconds) passively watching the game (in front of the TV but eyes not focused on the screen), and $y_{it} = 2$ if the agent spends > 30 seconds of minute t being actively attentive to the game (eyes focused on the screen).

The probability that agent i takes choice $j \in \{0, 1, 2\}$ at time t is specified as follows:

$$\begin{aligned}\Pr(y_{it} = j) &= \frac{\exp(\eta_{ijt})}{\sum_j \exp(\eta_{ijt})},\\ \eta_{ijt} &= f_j(z_t, SUS_t, SURP_t, CM_t, x_i).\end{aligned} \tag{4.2}$$

As to conditioning covariates, our key variables include the gameplay variables such as inning dummies and the absolute value of score difference (z_t), the suspense and surprise measures (SUS_t and $SURP_t$)[9], agent i's demographic variables (x_i), an indicator for whether a commercial is playing during minute t (CM_t); various interactions among these elements are also included. The model also controls for the day of the week dummies, game dummies and post-season dummies. Game dummies capture, for example, the effect of star pitchers and the importance of a given game such as a wild card game.[10] Note that this model is not fully structural, as the determinants of choice each minute depend on the current gameplay features, which strictly speaking are unknown to agents unless they have been viewing the game during the previous minute.[11]

In Table 3 we consider a very rich specification in which game features are interacted with suspense and surprise and viewer demographics.[12] Suspense by itself en-

[9] SUS_t, $SURP_t$ and z_{jt} are defined up until time period (minute) t

[10] In order to control for viewers' team preferences, we also estimate the specification with viewer fixed effects and found the results almost identical. Moreover, we estimate the specification with game-level clustered standard errors. We find the main variables remain statistically significant, while some variables become less statistically significant. That is because there are not many clusters in our data set. Lastly, we estimate the model with state-dependence by adding y_{it-1}. While there is evidence of strong persistence, the qualitative results change little.

[11] We have also explored a more structural version of the model where we explicitly model the beliefs of agents when they have not been viewing the game. While this model is substantively more complicated, preliminary estimates show little qualitative differences relative to the simpler model reported here.

[12] The model also controls for the interactions between the dummy variables for each inning and suspense and surprise. The coefficient estimates are available from the authors upon request. The results show that the positive effect of suspense on attention becomes even stronger in later innings.

Table 3: Estimates of the Multinomial Choice Model

	Passive attention	Active attention
Suspense	2.411*	4.215*
	(0.387)	(0.764)
Age	0.00191*	0.00852*
	(0.000331)	(0.000640)
Suspense x Age	0.00933*	0.0146*
	(0.00388)	(0.00700)
Female	-0.272*	-0.164*
	(0.0156)	(0.0290)
Suspense x Female	-0.768*	-1.391*
	(0.137)	(0.248)
Surprise	-0.356	0.0486
	(0.401)	(0.767)
Surprise x Age	0.00512	-0.00547
	(0.00556)	(0.0105)
Surprise x Female	-0.595*	-0.667+
	(0.196)	(0.365)
CM	0.175*	-0.0564
	(0.0237)	(0.0481)
Suspense x CM	0.826*	0.925*
	(0.214)	(0.428)
Surprise x CM	-0.144	0.261
	(0.258)	(0.475)
Postseson	0.576*	0.248*
	(0.0367)	(0.0668)
Suspense x Postseason	1.013*	1.288*
	(0.165)	(0.311)
Surprise x Postseason	1.707*	2.922*
	(0.236)	(0.453)
Absolute score dif	-0.0357*	-0.00160
	(0.00476)	(0.00915)
CM x Female	0.0309	0.163*
	(0.0246)	(0.0501)
Postseason x Female	0.169*	0.247*
	(0.0188)	(0.0361)
CM x Postseason	-0.228*	-0.259*
	(0.0257)	(0.0522)
Constant	-4.061*	-5.560*
	(0.0481)	(0.0932)
Observations	1008131	

Note: The model includes the game, day-of-the-week and inning dummies. The standard errors are heteroskedasticity robust.

ters positively into the determinants of both passive and active attention, whereas surprise enters insignificantly. But key interactions of suspense and surprise are significant. In particular, women are less responsive to both surprise and suspense than men. This finding may be counter-intuitive because one might consider that female viewers watch or pay attention to a baseball game only when the game is interesting enough, or when the game has a lot of suspense and surprise. Our results show the opposite, but this echoes existing papers examining the impact of suspense and surprise in televised sporting events, which also found women to be less responsive to these features (Gantz and Wenner 1991; Bizzozero, Flepp, and Franck 2016).

Moreover, we see that suspense *enhances* both passive and active attention to commercials, suggesting that viewers who are "glued" to the screen during suspenseful junctures pay greater attention to commercials which air during these times.[13] Existing experimental results (Nelson, Meyvis, and Galak 2009; Meyvis and Nelson 2007; Wang and Calder 2006) have studied how viewers respond to commercials that interrupt TV programs, but the evidence here is among the first to examine this in a field setting. At the least, it suggests that certain game features spill over *positively* into ad engagement, and we will explore these implications when we consider optimal program redesign in our counterfactuals below.

We also see that while active attention is lower during commercials, as may be expected, passive attention instead increases – this indicatess that although viewers pay less attention during commercials, they neither walk away from the TV nor "zap" to another station during commercial breaks. Also, both active and passive attention increase for postseason games, but the increase is much larger for passive attention (+56.7%) than for active attention (+29.9%). Thus, if postseason ads were priced based on the rating of the games – which includes both passive and active viewers – they would have paid almost twice as much for the ads than if they had used the (arguably more accurate) "active attention" measure.

[13]Overall, commercial breaks during baseball games occur in the middle and end of innings, as the teams are changing sides, which tend to be moments which lower suspense and surprise.

Table 4: Counterfactual: Reshuffle Commercials

	Data		**High Suspense CM**	
	Game	CM	Game	CM
Passive Attention	5.01%	6.01%	5.04%	6.28%
Active Attention	1.30%	1.29%	1.30%	1.38%

Women are more likely to be attentive to television commercials, an interesting finding as commercials during sports telecasts are typically targeted towards men,[14] and suggests that advertising could profitably adjust this targeting strategy.

5 Counterfactual Exercises

Next, we use our estimation results to consider counterfactual program designs.[15] This is particularly relevant for the baseball telecasts that we study, as the sale of television rights make up an overwhelming portion of a baseball team's revenue, and baseball owners and officials have been actively engaged in redesigning baseball programs including changing the rules of baseball to increase viewers' engagement and attention.

5.1 Changing the placement of commercials

In the first counterfactual, we consider the placement of commercials within baseball telecasts. Our results above indicated that suspense and surprise in gameplay affect how much viewers pay attention to intervening commercials. However, commercial breaks during baseball telecasts typically occur in the middle and end of each inning, as the teams are changing positions. Here, we consider the effects on viewing and

[14]The top 10 brands in terms of the total length of commercials during the baseball games include beer and technology brands, such as Coca Cola, Suntory, NTT Docomo, Kirin, and Asahi.

[15]Since our empirical analyses include all viewers who watch given game at least one minute, the sample includes many viewers who watch baseball games just limited time. Hence, we think the extensive margin effect of the counterfactual simulations is limited.

attention if commercial breaks occurred during suspenseful junctures in the game.[16] Specifically, we ran a counterfactual in which we moved all of the commercial minutes to the most suspenseful moments in the game.

These counterfactual results are reported in Table 4. The most noteworthy results here lie in comparing the second line of results, which contain the probability that viewers pay attention to the game and to the commercials. Compared to the data benchmark, we see that reshuffling the commercials to the most suspenseful moments in the game increases the probabilities of both passive (from 6% to 6.28%) and active attention (from 1.29% to 1.39%). These results effectively illustrate the spillovers from program content to commercial arising from our results.

5.2 Changes in Baseball Game Rules

5.2.1 Shortening Games

One rule change considered by professional baseball leagues is to shorten the length of the game. Specifically, MLB has discussed the possibility of shortening games to seven innings,[17] This is the second counterfactual we consider.

We implemented this counterfactual of shortening the game in two ways. First, we simply truncated all the games and redefined the winner as the team which was leading at the end of the seventh inning. This approach, while simple, is problematic as we know that viewer engagement (both viewing and attention) tends to be larger in inning 8 and 9 (cf., Fig 2). Furthermore, there is evidence that baseball managers would hypothetically treat a game shortened to seven innings as tactically equivalent

[16]Such a counterfactual scenario may arise in the near future, as currently TV advertising is undergoing a digital transformation where a fraction of TV ad spend is transacted in real time, which allows advertisers to make ad purchase decisions by considering real-time program content features. See https://www.emarketer.com/content/how-digitization-affects-tv-ad-sellers for more details.

[17]MLB discussed several potential rule changes for the 2020 season, e.g., https://www.cbssports.com/mlb/news/10-rule-changes-mlb-could-test-during-the-shortened-2020-season-including-universal-dh-and-a-mercy-rule/. In fact, MLB has changed rules quite often. For instance, MLB has decided to require any starting or relief pitcher to pitch to a minimum of three batters.

Table 5: Counterfactual Outcomes

	Data		Counterfactual					
			Shortened Game				Mercy Rule	
			Innings 1-7		Innings 3-9			
	Game	CM	Game	CM	Game	CM	Game	CM
Passive Attention	5.09	5.91	4.69	5.77	5.39	6.10	5.15	6.06
Active Attention	1.33	1.25	1.17	1.24	1.35	1.27	1.35	1.30
Total (passive+active)	6.42	7.16	5.86	7.01	6.74	7.37	6.50	7.36

Note: The table report the choice probabilities in the data and in the coutnerfactual simulations. The unit is the percentage.

to starting the game in the third inning.[18] For that reason, we also computed a second version of this counterfactual where we only retained the game data from innings 3-9.

Table 5 reports both sets of results. As expected, we see that shortening the game to seven innings when we remove the 8th and 9th innings decreases attention – both passive and active – to both the game as well as commercials. For instance, passive attention to the game falls from 5.09% in the data to only 4.69% in the counterfactual. This occurs not only because we removed the innings with the highest viewing, but also presumably because managers strategically adjust their tactics to ensure a competitive and exciting finish. Indeed, when we examine the results to the second version of the counterfactual, where we shortened the games by removing innings 1 and 2 for each game, the results alter diametrically: in this case passive and active attention to both the game and commercials increase relative to the benchmark 9-inning game.

5.2.2 Mercy Rule

Another rule change that has been discussed by MLB is introducing a mercy rule. Mercy rules are currently in use in a number of baseball leagues (in some European countries, Cuba, and Korea) as well as international competitions (e.g., little league and Baseball World Cup). To implement this counterfactual rule change, we assume that each game

[18] See https://www.northjersey.com/story/sports/mlb/mets/2020/08/29/ny-mets-adjust-seven-inning-games-pitching-strategies/5653170002/.

ends if one team outscores the opponent by more than or equal to 5 runs at the end of the inning after the 7-th inning.

The rightmost column in Table 5 reports the results of the counterfactual viewing behavior. We find that both passive and active attention to the game significantly increased: on average 5.15% (1.35%) of consumers paid passive (active) attention to the game under the mercy rule, compared to 5.09% and 1.33% in the benchmark case. Moreover, there is an striking asymmetry, as the mercy rule has larger effects on attention to commercials, than to the game. Compared to the data, passive attention to commercials increases from 5.91% to 6.06%, while active attention rises from 1.25% to 1.3%.

We also quantify the economic value of the game re-design. Since advertising rates are driven by viewership, if we aggregate this result to the market level, it implies that mercy rules could result in a 2.8% (from 7.16% to 7.36%) increase in ad revenues, or in dollar terms, an increase from $0.84 billion (90 billion yen) to $0.86 billion (92.5 billion yen).[19] The results highlight an interesting asymmetry between two game rule changes; Shortening games cuts games regardless of game situations, while the mercy rules cuts only one-sided games, which are less interesting. Those effects spill over to advertising engagement.

6 Conclusion

In this paper, we investigate the impact of telecast contents on audience engagement and propose program design strategies to optimize engagement. By using a unique dataset that tracks viewer facial expressions at high-frequency intervals, we carefully differentiate two engagement measures, passive attention and active attention. In the context of professional baseball games, we study whether program features, such as

[19]We assume that 50% of the baseball revenue come from advertising. https://asia.nikkei.com/Business/Media-Entertainment/Japan-s-baseball-league-places-several-bets-on-its-future

suspense and surprise, significantly influence viewer engagement. Overall, we find that suspense has a significant impact on both passive and active attention, but surprise less so. During commercials, although "viewership", as defined by the traditional Nielsen rating measurement, does not change, because consumers do not walk away from the TV or "zap" to another station, they indeed pay less attention. Using the estimated model, we simulate several counterfactual scenarios. Interestingly, we find that reshuffling commercials to the most suspenseful moments in the game increases attention; shortening the game has ambiguous effects on attention, but a "mercy rule", which selectively terminates non-competitive matchups, has small effects on attention to the game, but larger positive impacts on attention to commercials.

Our results provide important managerial implications to different stakeholders. For baseball teams and game designers who aim to attract viewership and attention from baseball fans, our results suggest that increasing the suspense level in games is the key. For advertisers, we find that the proposed rule changes can be an effective way to increase viewing and attention to commercials.

More broadly, our study highlights the value of granular data on consumer attention, without which these managerial insights cannot be found. Technology is enabling us to collect relevant consumer attention measures and design better contents to improve consumer welfare and advertising marketplace efficiency.

References

ALBERT, J. (2003): *Teaching Statistics Using Baseball*. The Mathematical Association of America, Washington DC.

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

BIZZOZERO, P., R. FLEPP, AND E. FRANCK (2016): "The importance of suspense and surprise in entertainment demand: Evidence from Wimbledon," *Journal of Economic Behavior Organization*, 130, 47 – 63.

BORLAND, J., AND R. MACDONALD (2003): "Demand for sport," *Oxford review of economic policy*, 19(4), 478–502.

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

BURAIMO, B., D. FORREST, I. G. MCHALE, AND J. TENA (2020): "Unscripted drama: soccer audience response to suspense, surprise, and shock," *Economic Inquiry*, 58(2), 881–896.

DANAHER, P. J. (1995): "What happens to television ratings during commercial breaks?," *Journal of Advertising Research*, 35(1), 37–37.

DENG, Y., AND C. F. MELA (2018): "TV viewing and advertising targeting," *Journal of Marketing Research*, 55(1), 99–118.

ELY, J., A. FRANKEL, AND E. KAMENICA (2015): "Suspense and Surprise," *Journal of Political Economy*, 123(1), 215–260.

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

GOETTLER, R. L., AND R. SHACHAR (2001): "Spatial competition in the network television industry," *RAND Journal of Economics*, pp. 624–656.

GUNTER, B., A. FURNHAM, AND Z. LINETON (1995): "Watching People Watching Television: what goes on in front of the TV set?," *Journal of Educational Television*, 21(3), 165–191.

HUTCHINSON, J., J. LU, AND E. WEINGARTEN (2017): "Visual attention in consumer settings.," in *Routledge international handbook of consumer psychology*, ed. by C. V. Jansson-Boyd, and M. J. Zawisza. Routledge/Taylor & Francis Group.

ITTI, L., AND P. BALDI (2009): "Bayesian surprise attracts human attention," *Vision research*, 49(10), 1295–1306.

KAPLAN, S. (2020): "Entertainment Utility from Suspense, Surprise, and Starpower," *mimeo*.

KNOEPFLE, D. T., J. T.-Y. WANG, AND C. F. CAMERER (2009): "Studying learning in games using eye-tracking," *Journal of the European Economic Association*, 7(2-3), 388–398.

MCGRANAGHAN, M., J. LIAUKONYTE, AND K. WILBUR (2019): "Watching People Watch TV," *mimeo*.

MEYVIS, T., AND L. NELSON (2007): "Contrasting against the future: The unexpected effects of expectation," *ACR North American Advances*.

NELSON, L. D., T. MEYVIS, AND J. GALAK (2009): "Enhancing the television-viewing experience through commercial interruptions," *Journal of Consumer Research*, 36(2), 160–172.

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

ROTTENBERG, S. (1956): "The baseball players' labor market," *Journal of political economy*, 64(3), 242–258.

SCHWEIDEL, D. A., AND R. J. KENT (2010): "Predictors of the gap between program and commercial audiences: An investigation using live tuning data," *Journal of Marketing*, 74(3), 18–33.

SIMONOV, A., R. URSU, AND C. ZHENG (2021): "Do Suspense and Surprise Drive Entertainment Demand? Evidence from Twitch.tv," *mimeo*.

SU-LIN, G., C. A. TUGGLE, M. A. MITROOK, S. H. COUSSEMENT, AND D. ZILLMANN (1997): "The thrill of a close game: Who enjoys it and who doesn't?," *Journal of Sport and Social Issues*, 21(1), 53–64.

TEIXEIRA, T. S., M. WEDEL, AND R. PIETERS (2010): "Moment-to-moment optimal branding in TV commercials: Preventing avoidance by pulsing," *Marketing Science*, 29(5), 783–804.

WANG, J., AND B. J. CALDER (2006): "Media transportation and advertising," *Journal of Consumer Research*, 33(2), 151–162.

WANG, J. T.-Y., M. SPEZIO, AND C. F. CAMERER (2010): "Pinocchio's pupil: using eyetracking and pupil dilation to understand truth telling and deception in sender-receiver games," *American Economic Review*, 100(3), 984–1007.

WILBUR, K. C. (2008): "How the digital video recorder (DVR) changes traditional television advertising," *Journal of Advertising*, 37(1), 143–149.

YAO, S., W. WENBO, AND Y. CHEN (2017): "TV Channel Search and Commercial Breaks," *Journal of Marketing Research*, 54(5), 671–686.

Appendix A Details for computing suspense and surprise

Following the standard approach used in the operations research literature (e.g., Albert, 2003), for the state transition, we can factor as follows.

$$\begin{aligned}
P(S_{t+1}|S_t) =& P(B/O', I', H', SD'|B/O, I, H, SD) \\
=& P(I', H'|B/O', SD', B/O, I, H, SD) * P(SD'|B/O', B/O, I, H, SD) \\
& * P(B/O'|B/O, I, H, SD) \\
=& P(I', H'|B/O', B/O, I, H) * P(SD'|B/O', B/O, SD) * P(B/O'|B/O, I, H, SD),
\end{aligned}$$

where I indicates an inning, H indicates whether top or bottom, SD indicates the score difference, B/O is the pair of bases loaded and out counts. Since there are eight possible states for which bases are loaded and three possible out counts, B/O can take 24 different possible values. The next period state is denoted with the prime $'$. The final equality just removes irrelevant conditioning variables from each term. We can estimate these three terms separately. Since the top of a innings is always the away team's offense and the bottom is the home team's offence, the first term is deterministic given the conditioning variables:

$$(I', H') = \begin{cases} (I, 1) & \text{if } H = 0, o = 2, o' = 0 \\ (I+1, 0) & \text{if } H = 1, o = 2, o' = 0, \end{cases} \tag{A.1}$$

where $H = 0$ means the top, $o = 2$ means the current out count is two.

The second term $(SD'|B/O', B/O, SD)$ is also deterministic, as for each transition of B/O from state i to state j in inning I, the associated run difference r_{ij}, equals:

$$r_{ij} = 1 + (b_i - b_j) - (o_j - o_i)$$

where b_i denotes the number of batters on base when B/O is in state i, and o_i denotes the number of outs in state i. Intuitively, the number of runs scored between states i and j equals the difference in active runners between states i and j minus the difference in outs between states j and i. Given r_{ij}, the score difference is

$$SD' = SD + H' * r_{B/O,B/O'} - (1-H') * r_{B/O,B/O'} \tag{A.2}$$

Lastly, for the third term $(B/O'|B/O, I, H, SD)$, we can estimate the 24×24 transition matrix for B/O separately for each value of (I, H, SD). Many elements in the transition matrix will be zero; for instance, it's impossible to go from $(\varnothing, 0)$ to $(12, 1)$.

Now we estimate the transition probability $P(S_{t+1}|S_t)$. We simply estimate its probability distribution function (pdf) with

$$\hat{p}(S_{t+1} = s'|S_t = s) = \frac{\sum_s \sum_{s'} \mathbb{1}\{S_t = s \; \& \; S_{t+1} = s'\}}{\sum_s \mathbb{1}\{S_t = s\}}, \tag{A.3}$$

where $\hat{p}(s'|s)$ is the pdf of $\hat{p}(\cdot)$, $s \in \mathscr{S}$ and $s' \in \mathscr{S}$. Note that the transition probability is deterministic from some states to others.

We directly estimate $P(hwin|S_t)$, the conditional probability that the home team wins, and $P(S_{t+1}|S_t)$, the transition matrix of states from the data, using the data on all professional baseball games in 2018, including the ones that are broadcasted. We first non-parametrically estimate $P(hwin|S)$ with XGBoost , a standard machine learning method. To avoid overfitting, we use a 5-fold cross validation.[20] Our model can correctly classify the home team's winning with 79.7% of accuracy. We explain the details of the estimation of $P(S_{t+1}|S_t)$ in the appendix.

With the estimates of those objects, $\hat{\mu}_t$ and $\hat{P}(s'|s)$, we can compute the expectation of the belief given the current state. For instance, the suspense measure can be computed as (let (i, j) denote the state of B/O this period and next period, respectively):

[20]We also tried Random Forest, Decision Tree, and Support Vector Machine to predict μ_t. We find that XGBoost outperforms other methods.

$$
\begin{aligned}
SUS_t &= E_t\left[(\mu_{t+1}-\mu_t)^2\right]^{1/2} \\
&= E_t\left[\left(\sum_{j=1}^{24}\hat{P}(hwin|j,I'_{ij},H'_{ij},SD'_{ij})\hat{P}(B/O'=j|B/O=i,I,H,SD)-\mu_t\right)^2\right]^{1/2},
\end{aligned}
\tag{A.4}
$$

where I'_{ij}, H'_{ij}, and SD'_{ij} denote the period-ahead values of I, H, SD.