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

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 gameplay 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 by 25%, but they walk away from the TV or "zap" to another station only by 8%. These results have implications for content design. While a "mercy rule" (which selectively shortens less competitive games) has small effects on attention, shortening baseball games increases attention for both games and commercials. Moreover, game redesigns improve attention of female viewers more than male viewers.

1 Introduction

Entertainment plays an important role in our life. An average American adult spends about 3 hours or 20% of their waking time, 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 watch the games on TV.[2]

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

[2] https://www.nielsen.com/us/en/insights/article/2020/during-covid-19-sports-viewers-are-still-a-

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, both suspense and surprise have positive and significant effects on passive and active attention, but suspense has a bigger effect on attention. This corroborates findings from the previous literature. Moreover, female viewers respond to suspense less than male viewers, while female viewers respond more to surprise than male viewers. Lastly, while both passive and active attention decrease during commercials, active attention decreases much more than passive attention. These findings have implications for program designs and commercial targeting.

Using these results, we simulate several counterfactual scenarios to assess the impact of rule changes in baseball, as the largest component of baseball revenue derives from television. First, shortening baseball games has positive effects on active attention to games and commercials, while it has a slightly negative effect on passive attention to games and a sizable positive effect on passive attention to commercials. Second, we find that a "mercy rule" which selectively shortens less competitive games (i.e., games where one team outscores the opponent by at least five runs after the 7th inning), has little effects on attention. This is probably because mercy rules are not applied to many games. Moreover, the simulation results indicate that the game redesigns improve the attention of female viewers more than male viewers. Thus, overall, we find that these proposed changes have important implications

scoring-opportunity-for-brands-and-media-owners/

for baseball teams and for advertisers. These 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 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, allowing us to distinguish whether viewers are passively or actively watching a program. As we will see, while viewing, 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

[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/

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

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.

Related Literature

Our paper is related to a number of streams of literature.

Measuring Attention: visual fixations. Using the AC Nielsen data (or similar set-top-box data) that measure whether people "tune" into a particular program, many studies have examined consumer program choices and switching decisions (e.g.,Rust and Alpert, 1984; Wilbur, 2008; Goettler and Shachar, 2001; Deng and Mela, 2018), as well as ad avoidance behaviors (e.g., Danaher, 1995; Schweidel and Kent, 2010; Teixeira, Wedel, and Pieters, 2010). Collectively, they identified viewer-related, program-related, and time-related factors that affect TV viewing behaviors.

However, this literature suffers from two limitations. One is that the rating metric can be a poor measurement of consumer engagement, as pointed out by Gunter, Furnham, and Lineton (1995) and Schmitt, Woolf, and Anderson (2003). The second limitation of this literature is that they focus on viewing choices across programs, instead of moment-to-moment choices within a program.

We address these two limitations in this paper. First of all, we leverage a novel dataset that directly measures 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 (2022), which studies a different research question about how contents in TV commercials affect viewability and attention. In contrast, we focus on the problem of program content design. Prior research has attempted to use VCRs and cameras to record viewers as

[5]See Hutchinson, Lu, and Weingarten (2017) for a study where eye-tracking data have been 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.

they watch TV (Ritson and Elliott 1999; Jayasinghe and Ritson 2013), an early predecessor to TVision's advanced technology. However, these studies are limited by their small sample sizes. Second, we construct factors that represent content dynamics within a program, and quantify their impact on consumer attention choices minute-by-minute.

Summarizing Program Content: Suspense and Surprise. Among all the 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; importantly, 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 can be rooted to the "outcome uncertainty hypothesis" in the sports economics literature (Rottenberg, 1956; Borland and MacDonald, 2003).

Several papers have examined empirically the relationship between suspense and surprise and TV rating in sports contexts. 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. An exception is a recent paper by Simonov, Ursu, and Zheng (2022), which looks at the impacts of suspense and surprise in e-sports viewing on twitch.tv using the viewer-level data. They also find that suspense is more important than surprise for viewing. Several key points differentiate our analysis from Simonov, Ursu, and Zheng (2022). First, we have access to the attention data, which allows us to measure consumer “engagement” (on top of viewing) to contents. Second, our content data includes both TV programs and commercials, which allow us to study the spillover effect of consumers attention for programs to commercials. Third, we leverage the rich set of demographics variables to compare gender differences of suspense and surprise.

2 Background and Data Sources

2.1 TV Attention Data

In this paper, a key contribution is that we utilize unique high-frequency data on viewers' *attention* to television programs and commercials. Our data come from TVISION INSIGHTS (hereafter TVISION), a company specializing in the collection of real-time television viewer response data utilizing eye-tracking software installed on viewers' television sets.

Specifically, we use the TVISION panel dataset (similar to Nielsen's TV panel), which consists of approximately 1,900 individuals in 800 households in Japan. Once a household joins the panel, TVISION installs a device with facial recognition capability to track each household member's TV viewing behavior. Those households receive certain payments by participating to the panel. The same dataset has been used in only one other marketing paper, McGranaghan, Liaukonyte, and Wilbur (2022), which studies a different research question about how contents in TV commercials affect viewability and attention.

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

Figure 1: Device and Facial Recognition System

Note: The images are provided by TVISION INSIGHTS.

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 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.[6]

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

[6]The facial recognition system also infers the emotion of each individual based on her/his facial expression as the right panel of Figure 1 shows. We do not have the emotion information for this study.

Table 1: Summary Statistics - Individuals' Viewing Behavior

	During each minute, % of viewers who:		
	Not tuned into baseball	Passive attention	Active attention
During gameplay minute	75.20%	19.24%	5.56%
During CM minute	78.18%	17.65%	4.17%
During all minutes	75.72%	18.96%	5.32%

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 two minutes of the game (because suspense and surprise are constructed based on the difference between two consecutive minutes). 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 years.

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" if she spends most of that minute (> 30 seconds) 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 (> 30 seconds) 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.[7]

[7]Note that these measures of attention are different from the raw data of attention that the company provided. Due to an NDA, we are not allowed to disclose the raw attention and viewing information. Note

Table 1 shows that viewers are more attentive to the game than to commercials (abbreviated "CM" in this paper) in terms of both passive and active attention, but active attention drops during commercials much more than passive attention (8.3% vs. 25%). This finding that active and passive attention move differently implies that viewers' attention choices may not be well captured by "single index" discrete choice models such as the ordered logit model or the probit model; rather, for our empirical work, we utilize the multinomial logit model, in which separate indices determine agents' choices to pay passive versus active attention.[8] Moreover, this difference suggests that the estimated engagement to commercials that is not based on active attention may be overestimated to the extent that active attention requires more viewer engagement.

2.2 Baseball Gameplay Data

Since our focus in this paper is on baseball telecasts, we supplement the attention data from TVISION with gameplay data on professional baseball matches in Japan. 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 reports summary statistics of all the games in 2018. Among 877 games in total, the home team wins a game with 49.4% probability, and scores about 4.28 points on average.

that if a person has 20 seconds not tuned to baseball, 20 seconds passive and 20 seconds active viewing in a one-minute period, then there is no "most of that minute" for any of these 3 activities. We checked the data and found no case like this. In the rare case when it happens, researchers can assign this minute randomly to the 3 activities.

[8] We estimate ordered logit models to check the robustness of the results and report the results in the appendix.

Table 2: Summary Statistics-Baseball Games

	Mean	Sd	Min	Max	N
All					
Home team win	.494	-	0	1	877
Home team score	4.279	3.079	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
Broadcasted only					
Home team win	.659	-	0	1	41
Home team score	4.512	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: A tie is not included in the event of home team winning the game. There are 17 games ended up with a tie at the end of 12th inning.

There are about 300 pitches in each game and each game lasts on average 12,000 seconds.[9] 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 65.9% probability and scores on average 4.5 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,[10] while other characteristics are similar among games. As the main television networks in Japan are headquartered in Tokyo, the broadcasted games tend to feature Tokyo-area teams, such as the Tokyo Giants and Yakurt Swallows (indeed, games involving the Tokyo-area teams make up 61% of the broadcasted games,[11] but are only 5% of all games). Moreover, due to their television presence, Tokyo-area teams tend to have a national fan base.

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 viewing and attention are slightly

[9] A game ends up in a tie if both teams run the same scores at the end of 12th inning. In the full data, there are 17 tie games (1.94%), while in the broadcasted games, there is only one tie game (2.4%).

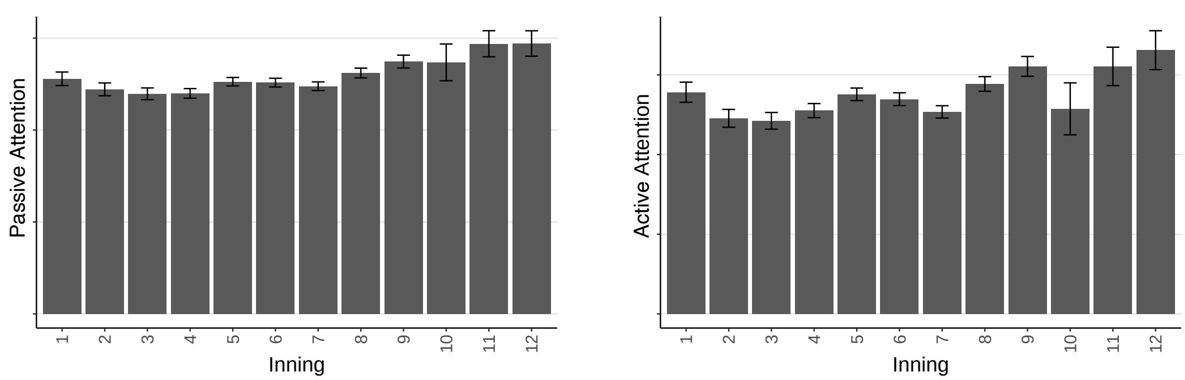
[10] This could happen because those games of top-seed teams when playing home are more likely to be broadcasted. We acknowledge that there could be a selection issue and this is a data limitation.

[11] The Tokyo-area teams are always playing home games when being broadcasted.

higher in the first inning.[12] This pattern indicates that people zap away from baseball games and switch to other programs as the game progresses and come back to later innings to check the outcomes of the game.

In Figure 3, we plot viewing (Figure 3 (a)) and attention (Figure 3 (b)) against the absolute value of the score difference. We find that viewing first decreases as the absolute score difference increases, while increases as the absolute score difference is higher than 4. Since the score difference tends to be less than 3 in many games, people tend to view the game when the score difference is small. On the other hand, attention increase when the score difference between the teams is smaller, indicating that viewers are more engaged when the game is more competitive. The contrast between viewing and attention in Figure 3 highlights the importance of collecting attention data beyond the rating/viewing data. Specifically, when the score difference is large, although consumers still keep the TV on, they no longer pay attention to the TV, so launching commercials during this time may be ineffective.

Figure 2: Viewing and Attention by Inning



(a) Passive Attention

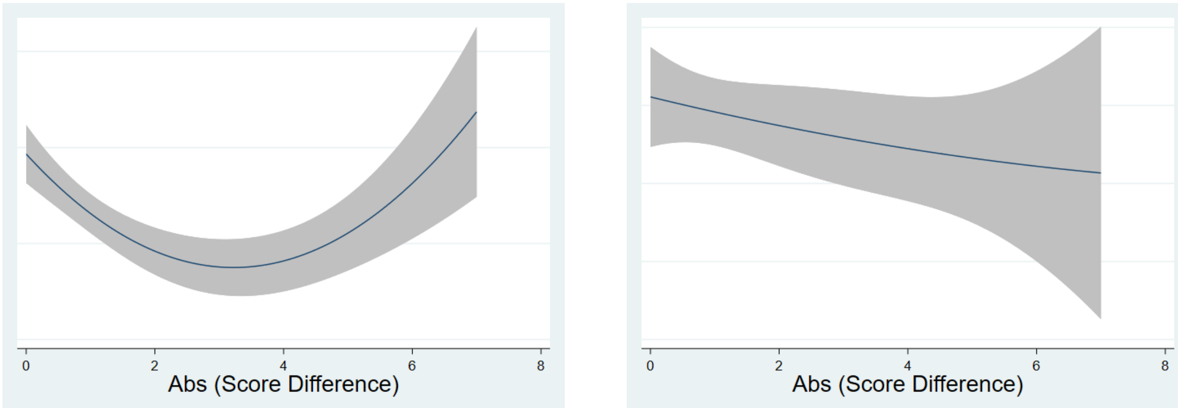
(b) Active Attention

Note: The figure plots the average passive attention (sitting in front of the TV but not paying attention) and active attention over innings. 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.

:

[12]Note that viewing and attention in Figure 2 are based on the raw measures of viewing and attention provided by the company, and hence they are different from passive and active attention measures in the previous section.

Figure 3: Viewing and Attention by Score Differences



(a) Viewing (b) Attention

Note: The figure plots the average viewing (sitting in front of the TV but not paying attention) and attention (paying active attention) over the score difference. 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. Even for a relatively simple game such as baseball, effectively and parsimoniously summarizing the gameplay features which are likely to influence viewer attention is not easy, as some game features may only be noteworthy if these occur in tandem (e.g., a runner on second base may merit more attention when there are currently two outs) or in a nonlinear fashion. Moreover, we also want to consider game features which are generalizable to other program genres, so that our results may have broader relevance beyond baseball.

For these reasons, we follow the recent literature and focus on two key determinants, suspense and surprise, that 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. For example, a suspenseful moment is the situation where a team has all bases loaded and is about to score, while a surprising moment is the situation where a team scores by a home run when no runner is on any base.

In the remainder of this section we provide technical descriptions on the construction of these variables; readers interested in our empirical model can skip ahead to Section 4.Ad

The construction of suspense and surprise proceed in two steps. In the first step, we estimate the viewer's belief of the game outcome, i.e., whether the home team wins. Also, we estimate the state transition probabilities directly from the data. Then, we calculate suspense and surprise using the estimated belief and state transition probabilities by applying the definition of suspense and surprise in Ely, Frankel, and Kamenica (2015).

Step 1: Calculate Winning Probability We assume that as they watch the game, viewers form and update beliefs regarding which team will win the game, i.e., the finite state space $\Omega = \{\text{home win}, \text{home lose}\}$.[13] We define that home win $= 1$ if the final score of the home team is greater than the final score of the away team. Hence, a tie game is classified as home lose in our definition. Our results are robust to an alternative definition of the state space with a tie. More specifically, at "at-bat" t in the baseball game, we let $\mu_t \in \Delta(\Omega)$ denote a typical belief, which is the perceived probability that the home team will win the game (this event is denoted "$hwin$"). We assume that viewers are rational and hold correct beliefs about the probability of the realization of the state. In other words, a belief represents the viewer's evolving expectations about which team will win the game given the information she gets.

We assume that belief $\mu_t \equiv P(hwin|S_t)$ can depend on state variables S_t of the game. Moreover, we assume that the belief μ_t follows a Markov process, i.e., the state transition probability $P(S_{t+1}|S_t)$ is Markov. These assumptions follow the existing literature.

Note that at each value of the state variables S_t, the probability $\hat{P}(hwin|S_t)$, is directly

[13] One would think about different sets of potential outcomes such as points in each inning by each team. We follow the existing literature which also define the state space as the final outcomes of the game.

Table 3: Confusion Matrix

Model/Data	0 (home lose)	1 (home win)
0 (home lose)	5503 (40.8%)	2126 (15.8%)
1 (home win)	1470 (10.9%)	4384 (32.5%)

The table reports the confusion matrix with the test data, which is 20% random sample. The horizontal axis shows the classification based on the model prediction, while the vertical axis shows the classification based on the observed data. The numbers in parentheses are the fraction of each case.

estimated from the data. Specifically, using the data on all baseball games in 2018, we non-parametrically estimate $P(hwin|S)$ with XGBoost (Chen and Guestrin, 2016), which is a standard machine learning method. To avoid overfitting, we use a 5-fold cross validation for tuning hyperparameters.[14]

The state variable is denoted by $S_t = (I, H, SD, B/O)$, where I indicates an inning, H indicates whether top or bottom, SD indicates the score difference, B/O is the pair of bases occupied 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 $'$.

Our construction of the state variables is based on the large literature of the baseball analytics in Operations Research (e.g., Albert, 2003). Although there are other variables that may affect the home team winning probability such as batting order and star players, we keep the state variables as parsimonious as possible and add other variables in the choice model later.

Table 3 shows the confusion matrix of our predictive model. Our model can correctly classify the home team's winning with 73.3% accuracy, based on the 20% random holdout sample. Also, the precision of our model is 74.9%, which is sufficiently high compared to win probabilities implied by betting odds of around 60% for experts and users on the

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

leaderboard.[15]

Step 2: Calculate Suspense and Surprise Following Ely, Frankel, and Kamenica (2015) and Bizzozero, Flepp, and Franck (2016), we define the suspense measure as

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

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:[16]

$$SURP_t = |\mu_t - \mu_{t-1}|. \tag{2}$$

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.

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

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

[15] Source: https://www.bettingpros.com/mlb/accuracy/game/moneyline/2022/0/

[16] 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.

The first equality decomposes the state transition into (i) the transition of the inning and the top/bottom in each inning, (ii) the transition of score differences, and (iii) the transition of bases loaded and out counts. 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, $P(I', H'|B/O', B/O, I, H)$, is deterministic given the conditioning variables:

$$(I', H') = \begin{cases} (I, 1) & \text{if } H = 0, O = 2, O' = 0, \\ (I+1, 0) & \text{if } H = 1, O = 2, O' = 0, \\ (I, H) & \text{otherwise,} \end{cases} \tag{3}$$

where $H = 0$ means the top, $O = 2$ means the current out count is two. Thus, the state (I, H) moves to another state only when the out count is two and the next state includes either next inning or the bottom of the inning.

The second term, $(SD'|B/O', B/O, SD)$, is also deterministic. To see this, note that, for each transition of B/O from state i to state j in inning I, the associated run difference r_{ij}, is given by the formula

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

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 is equal to the difference in active runners between states i and j minus the difference in outs between states j and i. Given r_{ij}, the corresponding score difference is

$$SD' = SD + H' * r_{B/O, B/O'} - (1 - H') * r_{B/O, B/O'}. \tag{4}$$

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, as it is not possible to move from certain states to others. 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

$$\hat{P}(S_{t+1} = s'|S_t = s) = \frac{\sum_s \sum_{s'} \mathbb{1}\{S_t = s \ \& \ S_{t+1} = s'\}}{\sum_s \mathbb{1}\{S_t = s\}}. \tag{5}$$

Note that the transition probability is deterministic from some states to others as discussed above.

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. The suspense measure can be computed as (let (i, j) denote the state of B/O this period and next period, respectively):

$$\begin{aligned} SUS_t &= E_t \left[(\mu_{t+1} - \mu_t)^2\right]^{1/2} \\ &= E_t \left[\left(\sum_{j=1}^{24} \hat{P}(hwin|j, I'_{ij}, H'_{ij}, SD'_{ij})\hat{P}(B/O' = j|B/O = i, I, H, SD) - \hat{\mu}_t\right)^2\right]^{1/2}, \end{aligned} \tag{6}$$

where I'_{ij}, H'_{ij}, and SD'_{ij} denote the period-ahead values of I, H, SD. Similarly, the surprise measure is computed by

$$SURP_t = |\hat{\mu}_t - \hat{\mu}_{t-1}|. \tag{7}$$

Summary Statistics of Suspense and Surprise Table 4 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 to compute suspense and surprise. The average suspense is 0.037, 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.[17]

[17] 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.

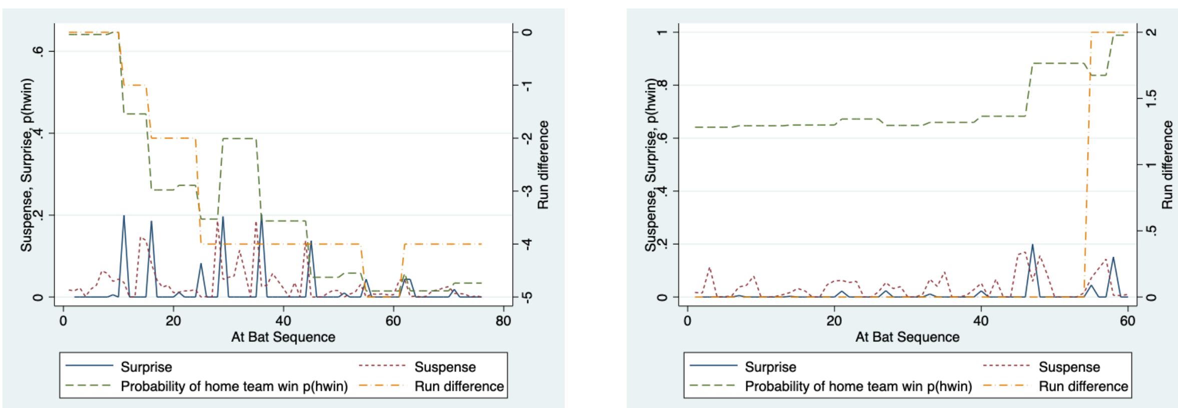
Table 4: Summary Statistics: Suspense and Surprise

	Suspense	Surprise
Mean	0.037	0.017
Std. dev.	0.062	0.069
Median	0.017	0
Obs.	67,411	66,534

Note: The suspense and surprise metrics are estimated with all baseball games in 2018.

Also, the median of surprise is zero, which implies that the distribution of surprise is skewed toward 0.

Figure 4: Examples of Suspense and Surprise Over Time



(a) Game 1

(b) Game 2

Note: Run difference is the difference between the home and away teams.

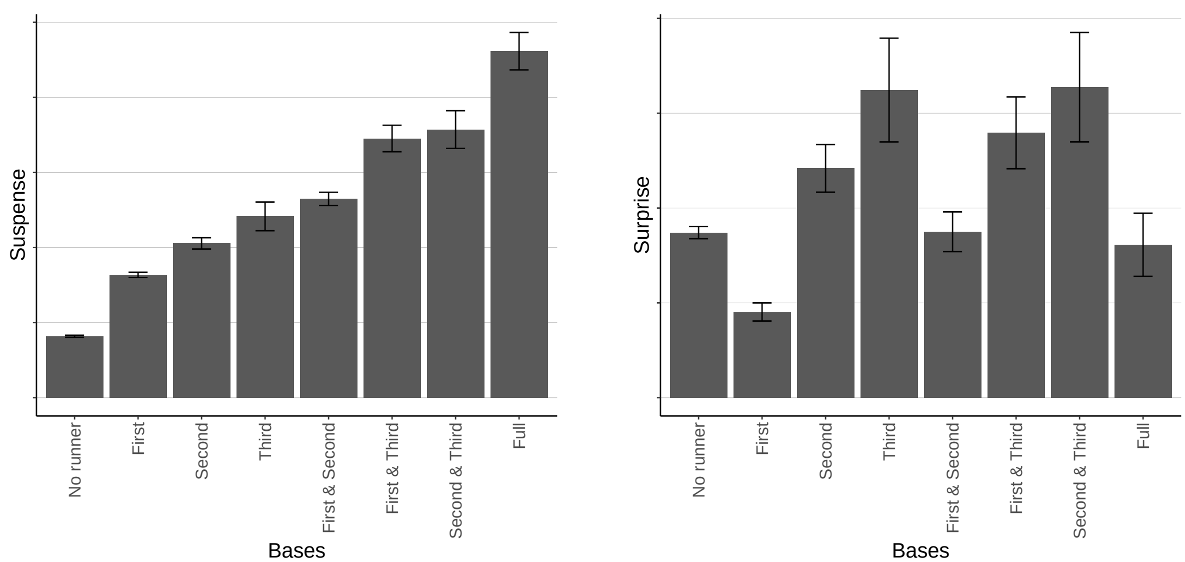
Two subfigures in Figure 4 show how the two measures (suspense and surprise) move within a game for two exemplar games. In the first game (panel (a)), the away team scores two points in the first half of the game, then adds more points later in the game. Hence the probability that the home team wins the game remains very small during the entire game, leading to very little suspense and surprise.[18]

By contrast, the game shown in panel (b) of Figure 4 is a more exciting one. The game is

[18]The plot is consistent with the intuition that there should be a trend in the measures of suspense and surprise as the game progresses. For instance, as the game gets closer to the end, it gets harder and harder for a losing team to recover, so suspense decreases. For this particular game, suspense became almost zero after the at bat sequence is greater than 30.

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. As those two examples illustrate, our suspense and surprise measures are closely related to what we usually think as suspenseful and surprising moments.

Figure 5: Suspense and Surprise by Whether Batters are On-Base



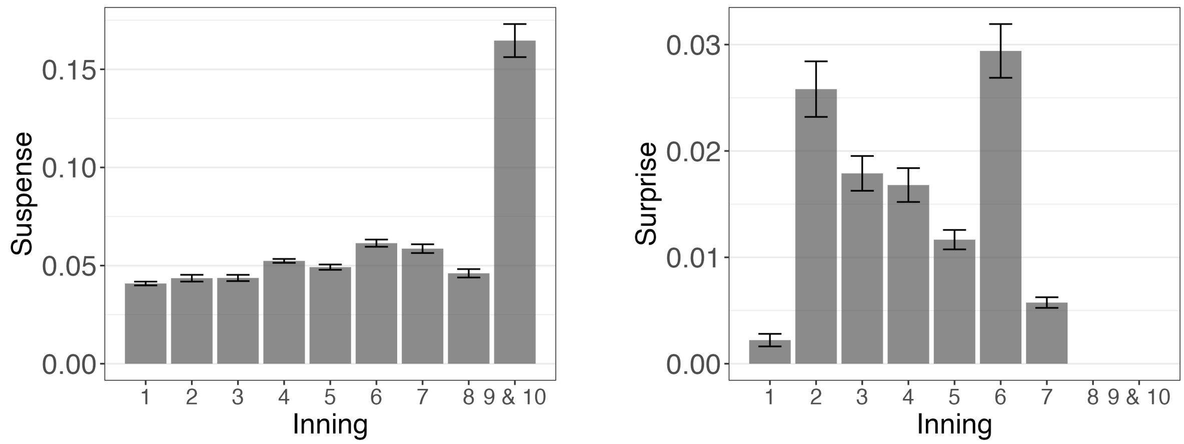
(a) Suspense (b) Surprise

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

Figure 6 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 mostly together with other game features. For example, as in Figure 7, the absolute difference in score between the competing teams is negatively correlated with both suspense and surprise (although surprise

Figure 6: Suspense and Surprise by Inning



(a) Suspense

(b) Surprise

Note: The figures plot the average and the 95% confidence intervals for suspense and surprise by inning.

increases a bit as the score difference increases from 0 to 2), implying that both suspense and surprise are greater when the game is a tie.

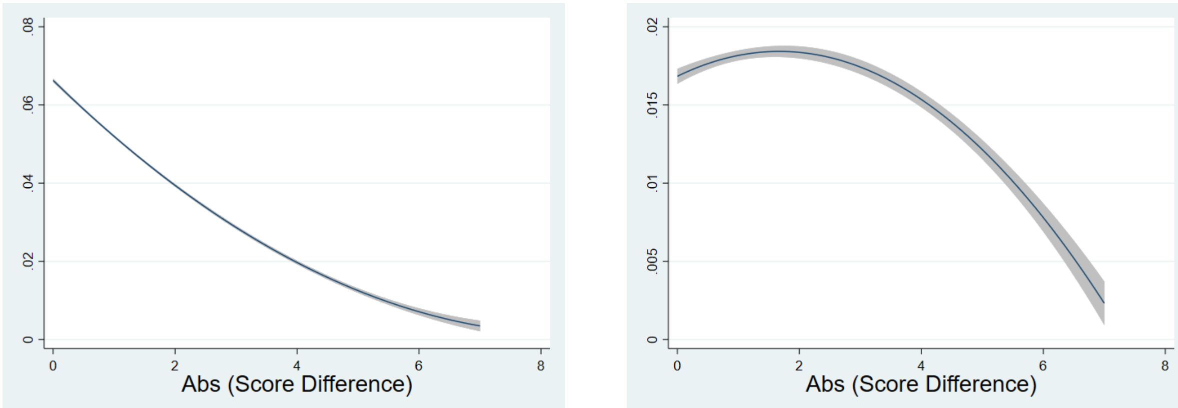
4 Viewing and Attention Choice Model

4.1 Viewer's Choice Model

For modeling agents' viewing and attention choices during the baseball game, we consider a multinomial logit model.[19] 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.

[19]A nested logit specification(Danaher and Dagger, 2012) generates qualitatively similar results. See the equation and regression results in Appendix XXX

Figure 7: Suspense and Surprise by Score Differences



(a) Suspense (b) Surprise

Note: Figures plot the average and the 95% confidence intervals for suspense and surprise over score differences.

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

$$y_{it} = \begin{cases} 0 & \text{if not tuned in to baseball telecast} \\ 1 & \text{if passively paying attention} \\ 2 & \text{if actively paying attention.} \end{cases} \tag{8}$$

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 agent's utility from watching TV depends on the gameplay variables (including suspense and surprise), commercials, and the agent's characteristics as follows

$$U_{ijt} = f_j(z_t, SUS_t, SURP_t, CM_t, x_{it}) + \varepsilon_{ijt}. \tag{9}$$

Note that Ely, Frankel, and Kamenica (2015) assume that the viewer's utility depends only on suspense or surprise. Also, other empirical papers studying the effects of suspense and surprise (e.g., Simonov, Ursu, and Zheng, 2022) examine the viewer preferences as a function of suspense and surprise. We assume that ε_{ijt} follows the logistic distribution. Hence, the choice probability at time t can be written as

$$\Pr(y_{it} = j) = \frac{\exp(f_j(z_t, SUS_t, SURP_t, CM_t, x_{it})}{\sum_j \exp(f_j(z_t, SUS_t, SURP_t, CM_t, x_{it}))}, \tag{10}$$

We assume that the consumers' utility is characterized by the following variables. First of all, the entertainment utility a consumer derives from watching a baseball game is captured by the content measures of suspense and surprise (SUS_t and $SURP_t$). SUS_t, $SURP_t$ and z_t are defined up until time period (minute) t.[20] Because people tend to change viewing decisions during commercial breaks, we add an indicator for whether a commercial is playing during minute t (CM_t) to the utility function. The levels of suspense and surprise when there is a commercial break remain the same as those in the last minute before the commercial break. Other gameplay features z_t includes inning fixed effects and game fixed effects. Lastly, we have an agent's demographic variable, x_{it}, including age and gender, and whether the consumer lives in Tokyo or not which indicates whether the consumer is a fan of the home team or not.

Moreover, a viewer's team preference strongly influences her utility to watch a particular game (i.e., a viewer may pay higher attention for her local team). However, because most of the games included in our sample are the Tokyo-area team games and the sample viewers are all fans of this team, there is no need to incorporate fan base in the utility function anymore. Furthermore, we account for the effect of star pitchers and the importance of a given game such as a wild card game using game dummies.[21]

[20] Note that although surprise is a function of what happened in the past, we are not assuming that historical events affects one's decision to watch going forward. Our assumption is that the surprise at the current moment t (which is the difference between μ_t and μ_{t-1}) affects the consumer's current decision at moment t.

[21] 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.

Note that this model is not fully dynamic, as the determinants of choice in 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.[22] Although those agents do not view current gameplay features including suspense and surprise, gameplay features may still influence agents' utility indirectly. For example, agents may receive the information from their friends or through other social network channels.[23] Hence, our preferred model is the multinomial logit model above. Also, (Ely, Frankel, and Kamenica, 2015) assumes that consumers are myopic and the current consumption utility depends only on the current level of suspense and surprise. Moreover, several recently published papers related to ours, including Simonov, Ursu, and Zheng (2022); Buraimo, Forrest, McHale, and Tena (2020); Bizzozero, Flepp, and Franck (2016) also assume that consumers are myopic.

4.1.1 Identification

Suspense and Surprise Our identification strategy relies on the path-independence assumption (Kaplan, 2020) that the current minute's game outcome is random conditional on the previous minutes' characteristics. In other words, the realizations of game events, which determine viewers' beliefs and resulting suspense and surprise, are stochastic. Both suspense and surprise are constructed based only on the pitch-level baseball information and hence those metrics are exogenous to viewing behavior. For instance, in the second inning, some games have a score of 1-1 (a high degree of suspense since the variance of beliefs of which team will win is high), while other games have a score of 0-5 (a low degree of suspense). We are interested in the effect of these in-game realizations on viewers' decisions, since only current stream viewers know the realized suspense and surprise levels. We include game fixed () effects to control for differences across games, such as the level of skill or fandom

[22] Say a viewer tune out at time t, what would be the SUS_{t+x} and $SURP_{t+x}$ when she returns to watch the game x minutes later?

[23] Although we have also explored a more structural version of the model where we explicitly model the dynamic evolution of the agent's belief, it is computationally prohibitive to estimate such a model at the minute-level (as gameplay features change) with a large number of fixed effects.

experienced by opposing teams in a game, specifics of the game (e.g. time of day of the broadcast), or the particular match between teams (e.g. two well-known teams playing together). We further include inning fixed effects () to control for particular stages of the game. The remaining variation in suspense and surprise measures is likely only due to a particular (random) score path realization. More formally, this setting leads to standard exclusion restrictions that allow us to identify viewers' tastes for suspense and surprise.

$$E(\epsilon_{ijt} SUS_t | z_t, CM_t, x_{it}) = 0$$
$$E(\epsilon_{ijt} SURP_t | z_t, CM_t, x_{it}) = 0$$

These moment conditions allow us to estimate fbss;bsrg from the OLS regression equation (7). Although we cannot directly test moment conditions (11)-(12), we can use data on viewers' join decisions to rule out other shocks to viewers' entertainment utility (e.g., advertising, word of mouth, stream rankings) that might be correlated with suspense and surprise realizations. If shocks to suspense and surprise correlate with particular time-of-day events or word-of-mouth utility shocks, the latter shocks should affect both joining and leaving decisions of the viewers, while suspense and surprise is known only to the current viewers of a stream. Thus, estimating the effect of suspense and surprise on viewers' join decision (equation 8) provides a useful placebo test, that should capture other correlated shocks to viewers entertainment utility, or any indirect effects of suspense and surprise on viewers' join decision. Formally, we use the following moment conditions to identify these effects

$$E(\epsilon^*_{ijt} SUS_t | z^*_t, CM^*_t, x^*_{it}) = 0$$
$$E(\epsilon^*_{ijt} SURP_t | z^*_t, CM^*_t, x^*_{it}) = 0$$

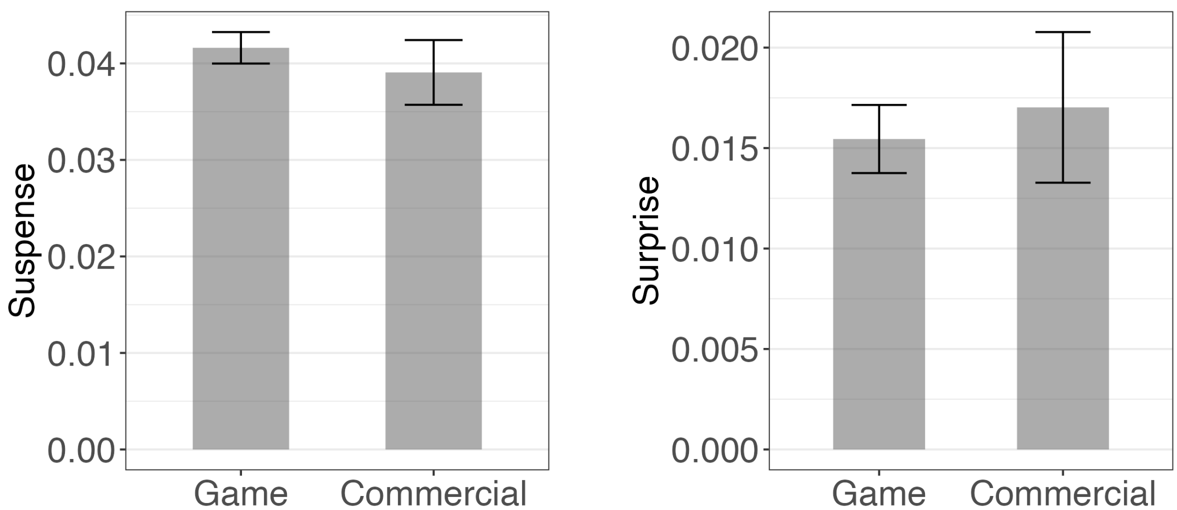
Null estimates of $\{\beta^*_{ss}, \beta^*_{sr}\}$ would rule out indirect effects of suspense and surprise on the viewers' decisions and serve as a placebo test supporting the validity of moments (11)-(12).

We cluster ϵ_{ijt} and ϵ^*_{ijt} at the game level since they might be correlated within a game.

Other game-play variables are also exogenous characteristics.

Commercial CM_t is also exogenous and it has little concern for endogeneity. This is because in Japan, timing of commercials is mostly predetermined between innings. For instance, Figure 8 shows that the suspense and surprise levels are not different for game minutes and commercial minutes, suggesting that commercials do not happen at more suspenseful or surprising moments of the games. Moreover, advertisers cannot target at game-minute level during baseball games, but rather they target at the game level. We include the game fixed effects, which capture such heterogeneity.

Figure 8: Suspense and Surprise by Game vs. Commercial



(a) Suspense (b) Surprise

Note: Figures plot the average and the 95% confidence intervals for suspense and surprise for the game minutes and commercial minutes.

4.2 Results

Table 5 contains estimates from this choice model. In the appendix, we report the estimation results of some robustness checks by estimating ordered logistic regression models. We start from the models without any interaction terms. The top panel reports the coefficients for the passive attention, and the bottom panel is for the active attention.

First, we find that both suspense and surprise positively affect passive attention and active attention, while the effect is bigger for suspense. This finding is consistent with other papers studying the effects of suspense and surprise (e.g., Simonov, Ursu, and Zheng, 2022 and Bizzozero, Flepp, and Franck, 2016), although those papers find insignificant effects of surprise. To interpret the magnitude of the viewers' preferences for suspense and surprise, we compute a change in the odds ratio of paying active attention to the game to paying passive attention to the game, $Pr(activeattentionatt)/(1 - Pr(activeattentionatt))$, driven by a change in the game's suspense. A one standard deviation increase of 0.062 in suspense leads to 1.229 * 0.062 = 0.0615 extra utils (taking model in Column 8 as the baseline), implying an increase in the odds ratio of exp(0:0615)-1 = 6.3%. Given that an average propensity of paying active attention to the game is 0.953, an odds ratio increase of 6.3% corresponds to 0.27 percent point increase in the probability of paying active attention.

Second, older viewers and male viewers are more attentive than young and female viewers. Third, both passive and active attention decrease significantly during commercials. Comparing the coefficients of commercials between passive and active attentions, we find active attention drops more during commercials than passive attention does. This indicates that during commercials, viewers do not pay attention to the screen, possibly by checking tablets and smartphones, even if they still tune in the baseball game.

In Table 6 we consider a very rich specification in which game features are interacted with suspense and surprise and viewer demographics. Suspense by itself enters positively into the determinants of both passive and active attention, whereas surprise enters insignificantly. In terms of the interaction terms, we find that women's active attention are less responsive to suspense than men, while women are more responsive to surprise than men. Interestingly, the results are not significant for passive attention. This finding could be intuitive because female viewers might be less interested in baseball and pay attention to a baseball game only

Table 5: Estimates of the Multinomial Choice Model

	(1)	(2)	(3)	(4)
		Passive Attention		
Suspense	0.189***	0.149*	0.0897	0.0865
	(0.0880)	(0.0901)	(0.0910)	(0.0940)
Surprise	0.0415	0.0946	0.0892	0.0251
	(0.0890)	(0.0913)	(0.0915)	(0.0934)
Age		0.0166***	0.0166***	0.0161***
		(0.000353)	(0.000354)	(0.000356)
Female		-0.176***	-0.176***	-0.173***
		(0.0120)	(0.0120)	(0.0120)
CM		-0.117***	-0.105***	-0.0844***
		(0.0158)	(0.0159)	(0.0161)
Constant	-1.394***	-1.565***	-1.584***	-1.115***
	(0.00724)	(0.0232)	(0.0337)	(0.152)
		Active Attention		
Suspense	0.638***	0.639***	0.614***	0.695***
	(0.153)	(0.151)	(0.153)	(0.159)
Surprise	0.289**	0.485***	0.451***	0.531***
	(0.153)	(0.151)	(0.149)	(0.156)
Age		0.0232***	0.0232***	0.0225***
		(0.000615)	(0.000617)	(0.000627)
Female		-0.0738***	-0.0733***	-0.0815***
		(0.0207)	(0.0207)	(0.0208)
CM		-0.316***	-0.296***	-0.303***
		(0.0295)	(0.0298)	(0.0301)
Constant	-2.690***	-2.997***	-3.043***	-4.892***
	(0.0128)	(0.0405)	(0.0597)	(0.717)
Day of week		X	X	X
Inning			X	X
Game				X
Obs	194808	194808	194808	194808

Note: The model includes the game, day-of-the-week and inning dummies. Standard errors are heteroskedasticity robust. *** indicates the coefficient is statistically significant at 1% level, *** indicates the coefficient is statistically significant at 5% level, and * indicates the coefficient is statistically significant at 10% level.

when some big events occur, such as a come-from-behind homer, rather than at suspenseful moments, which require continuous engagement to baseball games.

Table 6: Estimates of the Multinomial Choice Model with Interaction Terms

	(1)	
	Passive Attention	Active Attention
Suspense	0.554***	1.332***
	(0.174)	(0.300)
Age	0.0170***	0.0229***
	(0.000463)	(0.000799)
Suspense × Age	-0.0195***	-0.0119
	(0.00530)	(0.00875)
Female	-0.164***	-0.0767***
	(0.0158)	(0.0272)
Suspense × Female	-0.0683	-0.887***
	(0.180)	(0.300)
Surprise	0.0781	0.147
	(0.180)	(0.323)
Surprise × Age	-0.00693	0.00702
	(0.00538)	(0.00894)
Surprise × Female	0.219	0.641**
	(0.186)	(0.302)
CM	-0.0672**	-0.380***
	(0.0329)	(0.0658)
Suspense × CM	-0.0381	-0.206
	(0.272)	(0.493)
Surprise × CM	0.286	-0.487
	(0.239)	(0.465)
CM × Female	-0.0635**	0.188***
	(0.0324)	(0.0595)
CM × Age	0.0000375	0.000350
	(0.000945)	(0.00178)
Constant	-1.741***	-3.316***
	(0.0276)	(0.0497)
Day of Week	X	
Inning	X	
Game	X	
Obs	194808	

Note: The model includes the game, day-of-the-week and inning dummies. The standard errors are heteroskedasticity robust. *** indicates the coefficient is statistically significant at 1% level, *** indicates the coefficient is statistically significant at 5% level, and * indicates the coefficient is statistically significant at 10% level.

Table 7: Model Fit of the Multinomial Choice Model

	Choice Probability		
	$y = 0$ (Tune away)	$y = 1$ (Passive)	$y = 2$ (Active)
Data	75.20%	19.24%	5.56%
Model Prediction	75.19%	19.25%	5.57%

Although male viewers' passive and active attention decrease during commercials, female viewers' active attention increases during commercials. It is an interesting finding as commercials during sports telecasts are typically targeted towards men: indeed, the top advertisers during baseball games include alcohol and technology brands, such as Suntory, NTT Docomo, Kirin, and Asahi.[24] These findings suggest that advertisers could potentially adjust their targeting strategy in baseball telecasts and, in ongoing work, we are diving deeper to understand the ad categories which are draw particular attention from women. These gender interactions will also have important implications in the game redesign counterfactuals, which we describe below.

We also add interaction effects between suspense/surprise and commercial to capture the spillover effect from the game to the commercials.[25] The coefficients are insignificant, indicating limited spillover effect.

Model Fit Lastly, the multinomial logit model with interactions fits the data very well. Table 7 reports the choice probabilities in the data and the predicted choice probabilities (using coefficients in Table 6). To predict probabilities, we randomly split the sample into the training (80% of the data) and test samples (20% of the data). The prediction errors (i.e., percentage differences between the data and the prediction) are about 0.01%.

[24] We report the list of top 15 companies in terms of the length of commercials in the appendix Table A.4.

[25] The suspense and surprise during a commercial minute is the same as the minute before the game

Table 8: Counterfactual Suspense and Surprise

		Mean	Std. Err.	Obs
Data	Suspense	0.037	0.062	67411
	Surprise	0.017	0.069	66534
Mercy rule	Suspense	0.040	0.064	63601
	Surprise	0.019	0.072	62724
7 innings (3-9)	Suspense	0.042	0.072	51024
	Surprise	0.022	0.08	50147

Note: The table reports suspense and surprise in the data and in the counterfactual simulations.

5 Counterfactual Exercises: Changes in Baseball Game Rules

Next, we use our estimation results to consider counterfactual program designs.[26] In particular, we examine the effects of baseball game rule changes. Baseball rule changes are not without precedent. (MLB has, in fact, changed rules quite often.[27]) Such game redesigns are 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.

We focus on two potential rule changes which were discussed for the pandemic-shortened 2020 season:[28] (i) shortening games from 9 to seven innings; and a (ii) "mercy rules" which end the game when the score difference between the teams exceeds 5 runs by the end of the 7th innings. We consider these two redesigns in sequence.

[26]Since our empirical analyses include all viewers who watch given game at least two minutes, 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.

[27]https://www.baseball-almanac.com/rulechng.shtml

[28]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/.

Table 9: Counterfactual Rule Redesign: Shorten Games

(a) Attention Per Minute

	Data 7 innings (3-9)		CF 7 innings (3-9)		% Change	
	Game	CM	Game	CM	Game	CM
Passive Attention	19.07%	17.39%	19.04%	17.74%	-0.16%	2.01%
Male	20.06%	19.14%	20.01%	19.15%	-0.25%	0.05%
Female	17.73%	15.49%	17.74%	15.79%	0.06%	1.94%
Active Attention	5.46%	4.19%	5.60%	4.30%	2.56%	2.63%
Male	5.56%	4.03%	5.72%	4.06%	2.88%	0.74%
Female	5.41%	4.47%	5.44%	4.63%	0.55%	3.58%
Total Viewing	24.53%	21.58%	24.64%	22.04%	0.45%	2.13%
Male	25.62%	23.17%	25.73%	23.21%	0.43%	0.17%
Female	23.14%	19.96%	23.18%	20.42%	0.17%	2.30%

(b) Number of Minutes Paying Attention

	Data 7 innings (3-9)		CF 7 innings (3-9)		% Change	
	Game	CM	Game	CM	Game	CM
Passive Attention	19.07%	17.39%	19.04%	17.74%	-0.16%	2.01%
Male	20.06%	19.14%	20.01%	19.15%	-0.25%	0.05%
Female	17.73%	15.49%	17.74%	15.79%	0.06%	1.94%
Active Attention	5.46%	4.19%	5.60%	4.30%	2.56%	2.63%
Male	5.56%	4.03%	5.72%	4.06%	2.88%	0.74%
Female	5.41%	4.47%	5.44%	4.63%	0.55%	3.58%
Total Viewing	24.53%	21.58%	24.64%	22.04%	0.45%	2.13%
Male	25.62%	23.17%	25.73%	23.21%	0.43%	0.17%
Female	23.14%	19.96%	23.18%	20.42%	0.17%	2.30%

5.1 Shortening Games

One rule change which is under serious consideration by professional baseball leagues is to shorten the length of the game. Specifically, MLB has discussed the possibility of shortening games to seven innings. On the one hand, a shortened game could increase the suspense and surprise of games at each moment, therefore increasing consumer attention and effectiveness of commercials. On the other hand, a shorted game will reduce the number of commercial breaks and lead to less ad revenue. A priori, it is not clear whether shortening the game will make broadcasters and advertisers better off or worse off.

Note that implementing this counterfactual poses several challenges. Importantly, it is likely that baseball managers would alter their teams' strategies if games were shortened; they may ask batters to hit more aggressively, or ask pitchers to throw harder balls. However, it appears that, to a first order, managers may tactically treat a game shortened to seven innings as a nine-inning game which started in the third inning.[29] Following this insight, we implemented the counterfactual of shortening baseball games from nine to seven innings by retaining the game data from innings 3-9, and recomputing all the time-varying game features (in particular, the suspense and surprise measures) using gameplay in these innings.[30] We also note that in this counterfactual, the decisions of advertiser and TV networks (how many commercials to show) are assumed to remain unchanged.

The bottom row of Table 8 report the counterfactual suspense and surprise under the counterfactual scenarios. We found that both a mercy rule and shortening the game increases suspense and surprise. Comparing shortening games and the mercy rule, our simulations show that the shorter games in general lead to higher suspense and surprise. This is because the mercy rule may not be frequently applied. The number of observations in the mercy rule counterfactual is only 5.7% smaller than that in the original data, while that for shortening

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

[30] For completeness, we also ran the counterfactual using data from innings 1-7; the results, reported in Table A.3, are qualitatively quite similar to those reported in Table 9.

games is 24.3% smaller for suspense under 7 innings with dropping 1st and 2nd innings. Hence, even though both types of rule changes increase suspense and surprise, the impact of the mercy rule is smaller.

Table 9 reports the results of the counterfactual simulations under shortened games. We consider two metrics: one is the active/passive attention per minute, and the other is the total number of minutes paying active/passive attention. Note that the latter one is analogous to the Gross Rating Point (GRP) measure widely adopted in the advertising industry. Shortening games by retaining the later innings (3-9) increases passive and active attention, to both the game and to commercials compared to the data; however, in percentage terms, the increase in attention to commercials dwarfs the increase in attention to the game. For instance, in the shortening game counterfactual, passive attention to commercials increases by 2.01%, and active attention increases by 2.63%, while the increases in attention to the game are, in comparison, -0.16% and 2.56%. Clearly, game redesigns can have an appreciable impact on viewing and attention to advertisers.

Interestingly, there are sizeable gender differences[31] in these effects. Focusing on active attention, we see that the increase in active attention to the *game* is more prominent among men (+2.88%) than among women (only +0.55%). However, the active attention to the *commercials* is pronounced among women (+3.58%) but much less so among men (+0.74%). This result echoes the earlier findings from Table 6 that, compared to men, women are less likely to pay active attention to the game, but more likely to pay attention to the commercials, a tendency which is amplified in this counterfactual. x

Using these counterfactual results, we can also quantify the economic value of the game re-design. Since advertising rates are driven by viewership, and if we assume that 50% of the baseball revenue come from advertising,[32] then, at the market level, our results imply

[31] We report the gender differences for the population of baseball fans only. Differential selection into the types of consumers who watch sports across genders could exist. However, the selection will not bias our counterfactual results as long as we assume no change in the population considered in the counterfactual.

[32] https://asia.nikkei.com/Business/Media-Entertainment/Japan-s-baseball-league-places-several-bets-on-its-future

that shortening baseball games to seven innings would result in a 2.13%[33] (from 5.09% to 5.20%) increase in ad revenues, or in dollar terms, an increase from \$840 million (90 billion yen) to \$858 million (91.9 billion yen). Moreover, as shown in (McGranaghan, Liaukonyte, and Wilbur, 2022), attention (at 0.106) is significantly more indicative of ad effectiveness (measured by brand search lift) than if they were present in the room but not paying attention (0.013). Had advertisers considered giving a higher weight (0.106 vs 0.013) to the change in active attention than that in passive attention, then shortening baseball games to seven innings would result in a even larger revenue gain, to \$860 million[34] (92.2 billion yen). This exercise illustrates how program design may have important spillovers into advertising engagement.

5.2 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 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. Under the new rule, we develop a new predictive model of $P(hwin|S_t)$ and compute suspense and surprise using the new predicted values.

Table 10 reports the counterfactual outcomes for the mercy rule. The results show that both passive and active attention to the game and those to the commercials barely move. Even though the mercy rule could increase attention levels at a single moment due to greater suspense and surprise, it also decreases the total attention because the games become shorter, so the overall impacts of the mercy rule on the choice probabilities are small.

[33] 2.13% is the percentage change in total viewing. We assume that advertisers use the total viewing to set prices for ads.

[34] =840/(17.39%*0.013 + 4.19% * 0.106)*(17.74%*0.013 + 4.30% * 0.106)

Table 10: Counterfactual Rule Redesign: Mercy Rule

	Data		CF Mercy		% Change	
	Game	CM	Game	CM	Game	CM
Passive Attention	19.24%	17.65%	19.20%	17.65%	-0.21%	0.00%
Male	20.21%	19.06%	20.18%	19.04%	-0.15%	-0.10%
Female	17.91%	15.72%	17.88%	15.71%	-0.17%	-0.06%
Active Attention	5.56%	4.17%	5.57%	4.18%	0.18%	0.24%
Male	5.68%	3.94%	5.70%	3.95%	0.35%	0.25%
Female	5.40%	4.48%	5.40%	4.47%	0.00%	-0.22%
Total Viewing	24.80%	21.82%	24.77%	21.83%	-0.12%	0.05%
Male	25.89%	23.00%	25.88%	22.99%	-0.04%	-0.04%
Female	23.31%	20.20%	23.28%	20.18%	-0.13%	-0.10%

Note: In the first two columns, we report the choice probabilities in the data and in the counterfactual simulations. The unit is the percentage. In the third column, we report the change in percentage.

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 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(Van Meurs, 1998). Using the estimated model, we simulate several counterfactual scenarios. Interestingly, we find that shortening the game has positive effects on both attention and viewing, towards both the game and commercials, but a "mercy rule" has minimal impacts.

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.

A number of issues are left for future research. Our study discovers differential impact of suspense and surprise on passive and active attention. However, the underlying mechanism is not clearly understood. Future research could conduct experiments to better understand the underlying mechanism. Although this analysis of baseball telecasts is an important first step to understand the impact of program contents on engagement, more research on other types of TV programs are necessary. In dramas or quiz shows, for example, more work is required to define key measures of contents, analogously to the suspense and surprise measures used in this paper. Second, we have not yet examined the optimal timing of including commercials. Since commercials generally interrupt audience engagement, it is important to consider when to air commercials so that they interrupt telecasts in a way which maximizes audience engagement. Lastly, future research could link the price of the commercials to active versus passive attention and examine the broadcasters' optimal pricing strategies.

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 and 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.

Chen, T., and C. Guestrin (2016): "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794.

Danaher, P., and T. Dagger (2012): "Using a nested logit model to forecast television ratings," *International Journal of Forecasting*, 28(3), 607–622.

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.

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.

Jayasinghe, L., and M. Ritson (2013): "Everyday advertising context: An ethnography of advertising response in the family living room," *Journal of Consumer Research*, 40(1), 104–121.

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. C. Wilbur (2022): “How Viewer Tuning, Presence, and Attention Respond to Ad Content and Predict Brand Search Lift,” *Marketing Science*.

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.

Ritson, M., and R. Elliott (1999): “The social uses of advertising: an ethnographic study of adolescent advertising audiences,” *Journal of Consumer research*, 26(3), 260–277.

Rottenberg, S. (1956): “The baseball players’ labor market,” *Journal of political economy*, 64(3), 242–258.

Rust, R. T., and M. I. Alpert (1984): “An audience flow model of television viewing choice,” *Marketing Science*, 3(2), 113–124.

Schmitt, K. L., K. D. Woolf, and D. R. Anderson (2003): “Viewing the viewers: Viewing behaviors by children and adults during television programs and commercials,” *Journal of Communication*, 53(2), 265–281.

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 (2022): “Suspense and Surprise in Media Product Design: Evidence from Twitch. tv,” *Journal of Marketing Research, forthcoming.*

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.

Van Meurs, L. (1998): “Zapp! A study on switching behavior during commercial breaks,” *Journal of Advertising Research*, 38(1), 43–44.

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.

Appendix A Robustness Checks

A.1 Ordered Logit

In the main text, we estimate the viewer's preference with multinominal logistic model. In this section we consider a robustness check of which specifies the choice as an ordered choice. A rationale for this specification is that the viewer is more likely to pay attention to the TV screen as her satisfaction level increases, and less likely to watch the game if the satisfaction is low. Hence, the choice may be ordered. We estimate the ordered logistic regression model without interactions and with interactions and report the results below.

In Table A.1, we find that the results are consistent with the multinomial logit models. Both suspense and surprise affect positively to satisfaction. Also, older viewers and male viewers tend to watch baseball games more than younger and female viewers. Lastly, people watch less during commercials. Note that a drawback of the ordered logit model is that it is no longer possible to estimate the effects of program contents by low attention viewing and high attention viewing separately. Instead, we estimate only the combined effect of low and high attention viewing behavior. For example, in the multinomial logit model in Table 5 suspense and surprise enter passive attention insignificantly and active attention significantly. The ordered logit model cannot see the impacts separately.

Table A.2 reports the ordered logit results with interaction terms. Again, the results are mostly consistent with the multinomial regression. One exception is that in the multinomial regression, we find that the effect of Female$\times$CM for the passive attention viewing is negative, while that for the active attention viewing is positive. Combining the results, the ordered logit results tell that the total effect is negative but not statistically significant.

A.2 Additional Counterfactual Results

In the main text, we report the results of the counterfactual simulations for the shortened game when the first and second innings are dropped. We do so because of the anecdotal

Table A.1: Estimates of the Ordered Logit: Without heterogeneity

	Ordered Logit			
	(1)	(2)	(3)	(4)
Suspense	0.255***	0.293***	0.246***	0.265***
	(0.0798)	(0.0806)	(0.0821)	(0.0846)
Surprise	0.124	0.208**	0.193**	0.166**
	(0.0815)	(0.0823)	(0.0825)	(0.0842)
Age	0.0180***	0.0181***	0.0181***	0.0175***
	(0.000314)	(0.000314)	(0.000314)	(0.000318)
Female	-0.147***	-0.144***	-0.145***	-0.145***
	(0.0108)	(0.0108)	(0.0108)	(0.0108)
CM	-0.167***	-0.166***	-0.152***	-0.140***
	(0.0144)	(0.0144)	(0.0145)	(0.0147)
cut1	1.476***	1.499***	1.525***	1.422***
	(0.0112)	(0.0155)	(0.0269)	(0.151)
cut2	3.238***	3.262***	3.289***	3.191***
	(0.0143)	(0.0179)	(0.0283)	(0.151)
Day of week		X	X	X
Inning			X	X
Game				X
Observations	194808	194808	194808	194808

Note: The model includes the game, day-of-the-week and inning dummies.

stories by the professional baseball players and managers. In this section, we check the robustness of the results by dropping the eighth and ninth innings.

Table A.3 reports the simulation results. The results are both quantitatively and qualitatively similar to the results in the main text. We find that shortening games leads to an increase in both passive and active attention for games and commercials. In particular, the effect on commercial is greater than the effect on games. Also, similar to the results in the main text, attention by female viewers increases more than male viewers.

Table A.2: Estimates of the Ordered Logit: With heterogeneity

	(1) Ordered Logit
Suspense	0.794***
	(0.157)
Surprise	0.0753
	(0.163)
Age	0.0183***
	(0.000417)
Female	-0.135***
	(0.0143)
CM	-0.126***
	(0.0302)
Suspense x Age	-0.0156***
	(0.00474)
Suspense x Female	-0.325**
	(0.161)
Surprise x Age	-0.00232
	(0.00483)
Surprise x Female	0.350**
	(0.166)
Suspense x CM	-0.145
	(0.248)
Surprise x CM	0.0379
	(0.220)
Female x CM	-0.00654
	(0.0294)
Age x CM	-0.000224
	(0.000863)
cut1	1.444***
	(0.151)
cut2	3.213***
	(0.151)
Observations	194808

Note: The model includes the game, day-of-the-week and inning dummies. The model also includes the interaction terms between inning dummies and suspense and surprise, but the estimation results are not reported.

Appendix B List of Brands

In Table A.4, we list top 15 companies that have commercials during the baseball games in 2018. We exclude commercials that promote the programs of their own TV channel. The category is defined by TVISION. Out of 15 brands, 5 companies are in the beverage category,

Table A.3: Counterfactual Outcomes: Shortening Games (1-7 Innings)

	Data 7 innings (1-7)		CF 7 innings (1-7)		% Change	
	Game	CM	Game	CM	Game	CM
Passive Attention	18.87%	17.54%	18.90%	17.64%	0.16%	0.57%
Male	20.03%	19.23%	19.89%	19.06%	-0.70%	-0.88%
Female	17.31%	15.27%	17.57%	15.66%	1.50%	2.55%
Active Attention	5.28%	3.98%	5.46%	4.24%	3.41%	6.53%
Male	5.41%	3.78%	5.60%	4.02%	3.51%	6.35%
Female	5.11%	4.24%	5.26%	4.53%	2.94%	6.84%
Total Viewing	24.15%	21.52%	24.36%	21.88%	0.87%	1.67%
Male	25.44%	23.01%	25.49%	23.08%	0.20%	0.30%
Female	22.42%	19.51%	22.83%	20.19%	1.83%	3.49%

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

which sell beer and sport drinks. Hence, it seems that those brands target male viewers.

Table A.4: List of Companies with Commercials

Company	Category	Percent
Coca Cola	Beverage	4.08%
Suntory	Alcoholic beverage	2.68%
NTT Docomo	IT/Communication	1.59%
Recruit	HR	1.59%
Lion	Household products	1.49%
Nitori	Household products	1.29%
McDonald's	Food	1.29%
Kirin	Alcoholic beverage	1.29%
Asahi	Alcoholic beverage	1.19%
Amazon Japan	Retail	1.09%
GlaxoSmithKline	Pharma	0.99%
Suzuki	Automobile	0.99%
Toyota	Automobile	0.99%
Asahi	Beverage	0.80%

Note: The third column reports the share of each company among all commercials.