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

Xiao Liu* Matthew Shum† Kosuke Uetake‡

Abstract

Standard measures of television engagement, such as Nielsen-style "viewing," record whether a household is tuned in, not whether anyone is watching. We study how content and commercial breaks shape attention to baseball telecasts, using second-by-second data that log whether a viewer is not tuned in, passively present, or actively watching. Tuned-in viewers watch the screen only about 20 percent of the time. We model two linked choices—tuning in and watching actively—where the second is conditional on the first. Expected suspense raises tune-in, while realized surprise raises active attention. We use the results to inform changes in telecast design. First, repositioning commercial breaks toward the end of games, when gameplay drama is highest, raises active attention by about 2 percent without reducing attention to the ads. Second, shortening games to seven innings to enhance gameplay drama proves ineffective because the shorter telecast reduces attention more than the higher-stakes gameplay raises it.

*New York University, Stern School of Business. xliu@stern.nyu.edu
†California Institute of Technology and Amazon. mshum@caltech.edu
‡Yale School of Management. kosuke.uetake@yale.edu

1 Introduction

Television audiences are measured by viewing, which records whether a household's set is tuned to a program, usually through a set-top box, but not whether anyone is watching it. The gap between the two has grown as smartphones and tablets draw attention away from the set. In this paper we study how the content of a television program and the commercials within it shape audience engagement, and how the broadcast might be redesigned to raise it, using data that record not just whether a household is tuned in but whether each viewer is looking at the screen.

We use a unique dataset that logs, second by second, not only whether an individual is tuned in but also that individual's level of attention, so that we can separate *active* attention, when the viewer's eyes are on the screen, from *passive* attention, when the viewer sits in front of the set without watching. Conventional viewing measurement combines these two states, thus treating a distracted viewer the same as an engaged one.

This is especially problemmatic for sporting events, where moments of genuine excitement are randomly sprinkled within a generally routine viewing experience. Examining baseball telecasts during the 2018 Japanese professional baseball season, we find that the gap between passive and active viewing is large. Among viewers who are tuned in to the game, only about one in five is watching the screen at any moment; the rest are present but looking elsewhere. The gap is widest during commercials. Our results suggest a commercial lowers the odds that a viewer tunes in by about a third and the odds that a tuned-in viewer watches actively by about 30 percent. The audience that advertisers buy is thus both smaller and less attentive than the program audience, which neither conventional measures nor the raw attention shares reveal.

Based on this evidence, we develop an empirical model of viewing behavior during baseball telecasts, and recover parameters mapping viewers' decisions to telecast content. Every minute, the viewer makes two linked viewing decisions each minute: on the *extensive margin* whether to tune in to the game and, on the *intensive margin* conditional on being tuned in, whether to watch actively. Baseball forms a well-suited laboratory for our research. Baseball is very popular in Japan, and in 2019 (our study year)games drew an average of 30,929 spectators (against 28,317 for an average US Major League Baseball game), and they are aired frequently on television. Importantly for us, baseball rules are relatively simple so that game content is easily summarized by a low dimensional set of state variables (innings, outs, runners on base, and score differences). Finally, commercial breaks are scheduled around natural stoppages in play, which gives the broadcaster a concrete design lever whose consequences we can simulate.

Within this framework, we include content features which we expect to influence viewers' attention decisions. This include team features such as the identity of the competing teams, and batting order. To capture the dynamic and evolving stakes of a baseball game, we follow recent work and define *suspense* and *surprise*, formalized by Ely, Frankel, and Kamenica (2015). We find that the two operate on different margins. Expected suspense raises the probability of tuning in, while realized surprise raises the probability that a tuned-in viewer watches actively. The responses also differ by fandom: a one-standard-deviation increase in suspense raises active attention by about

1.6 percentage points among home-team fans against 0.3 points among neutral viewers, while a one-standard-deviation increase in surprise lowers active attention among away-team fans by about 3 percentage points.

Finally, we use the estimated model to ask how redesigning the broadcast reallocates attention between the game and the commercials that air within it, a question of direct interest to broadcasters as a large share of baseball revenue derives from television. We study the timing of commercial breaks: holding fixed each game's commercial blocks and total advertising time, we consider alternative placements of the blocks to maximize different objectives. Choosing break times to maximize total active attention raises it by $\tilde{2}.1\%$, without loss of attention to the commercials. Alternatively, the broadcaster can choose break times to maximize active attention to the commercials themselves by stacking more breaks during the later innings, which raises it by $+7.8\%$ without statistically significant decreases in attention to the game.

Related Literature

Our work departs from a large literature on television audience engagement in two ways that are worth stating at the outset. First, that literature has relied almost entirely on *viewing*, measured by set-top boxes in a panel of households, which is a coarse proxy for engagement because a household may leave the set tuned in while its members attend to something else, a concern that is especially acute during commercial breaks. Our attention data, collected via set-top sensors that detect faces and eye movements, record whether each viewer is actually watching, second by second. Second, prior work has largely studied engagement with programs while setting aside the commercials that air within them, even though those commercials are central to the business model of television. Our high-frequency data let us follow attention as it passes from the program into the surrounding breaks and back; because attention is persistent and viewers who look away lose track of the game, the attention lost at a break carries into the following minutes, so the timing of breaks matters for attention to both the commercials and the game. To our knowledge this interaction between program content, break timing, and attention has not been examined before.

Measuring television engagement. Using 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.[1] 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 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. This strand of researchmeasures engagement with aggregate TV ratings. These studies have reached conflicting conclusions about the relative importance of suspense and surprise: Bizzozero, Flepp, and Franck (2016) finds that surprise has a larger impact than suspense in Wimbledon tennis, whereas Kaplan (2020) (for basketball) and Buraimo, Forrest, McHale, and Tena (2020) (for soccer) find that surprise induces a smaller viewership response. Our findings that suspense impacts more viewers' tune-in decisions, and surprise impacts more the active attention decision, provide one possible resolution for these difference in results. A recent paper by Simonov, Ursu, and Zheng (2022) on e-sports viewing on twitch.tv using viewer-level data find that suspense is more important than surprise for viewing. We build on these analyses and results by exploiting our attention data, which allows us to measure consumer "engagement" (on top of viewing) to both the intrinsic game contents and to commercials.

Commercial breaks. A related literature studies audience behavior during commercial breaks using ratings or tuning data, documenting that audiences shrink and switch channels when breaks

[1] 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.

begin (Danaher, 1995; Van Meurs, 1998; Schweidel and Kent, 2010; Yao, Wang, and Chen, 2017), and how ad-skipping technologies change the demand for advertising (Wilbur, 2008; Tuchman, Nair, and Gardete, 2018). Teixeira, Wedel, and Pieters (2010) use eye tracking to study moment-to-moment avoidance within commercials. We add to this literature by modeling how the content of the program surrounding a break shapes attention, and by using the model to evaluate the timing of breaks.

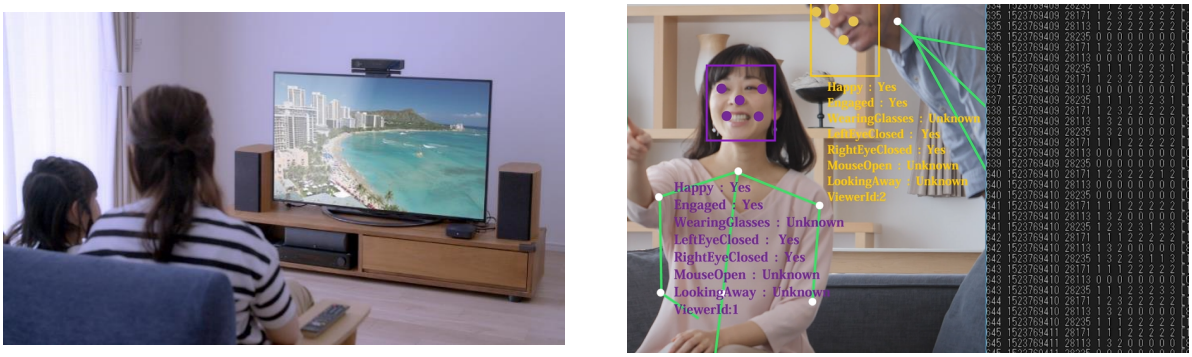
2 Background and Data Sources

2.1 TV Attention Data

In this paper, a key contribution our utilization of high-frequency data on viewers' *attention* to television programs and commercials. Our data come from TVISION INSIGHTS (hereafter TVISION),[2] a company specializing in the collection of real-time television viewer response data utilizing facial recognition 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 in the panel.

Figure 1: Device and Facial Recognition System



Note: The images are provided by TVISION INSIGHTS.

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.

The left picture in Figure 1 shows how the device is installed in each household. A key difference

[2]The company changed its name to REVISIO in 2023.

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.[3]

Based on this rich high-frequency data, we construct *three mutually exclusive states* of an individual's television engagement: (i) she is not watching the baseball telecast; (ii) she is tuned in and in front of the TV but engaged in another activity (*passive*); and (iii) she is tuned in and actively watching the screen (*active*). The typical "viewing" measure used by, for example, Nielsen, combines states (ii) and (iii), and does not distinguish between passive and active watching. In what follows, we use "tuned in" to denote conventional viewing, which includes both passive and attention attention, while we use our data to split attention into passive and active components.

Sample. Our data cover the 2018 season of Nippon Professional Baseball (NPB). The panel households are located in the Kanto region around Tokyo. We observe 46 televised games, 6 of them postseason games, and 1,157 panelists who watched baseball on the day of a game. 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, because events during a baseball game do not occur every second. Time is measured in minutes of game play, from the first pitch to the last out.

We classify each viewer-minute into one of the three states based on the viewer's action during most of the minute ("not tuned in" "active", "passive"). 26 percent of tuned-in minutes are active under the descriptive rule (Table 1).For estimation we keep each viewer's games in which she was tuned in for at least two minutes. This estimation sample contains 709 viewers, 2,374 viewer-game pairs, and 380,092 viewer-minutes; 45.6% of viewers are female and the average age is 40.6 years.

Commercial minutes. The device records the advertisements aired during each program. We define a minute as a commercial minute when the broadcast on screen shows more than 30 seconds of advertising. We construct this flag at the level of the broadcast, from all panel households watching it, so that it records when the broadcaster aired commercials rather than when a particular viewer happened to have the set on. Commercial minutes make up 11.6 percent of game minutes.

Table 1 reports the shares of viewer-minutes in each state. Two features stand out. First, most viewers of a game are not tuned in at any given minute, and among those who are tuned in, only about a quarter are actively watching the screen. A viewing-based metric, which counts every tuned-in minute alike, misses this gap between presence and attention. Second, the raw shares differ little between gameplay minutes and commercial minutes: the share not tuned in is 77.3 percent during gameplay and 77.7 percent during commercials, and the active share is 5.9 and 5.7 percent. This similarity does not mean that commercials leave attention unchanged. Viewers' states are highly persistent from one minute to the next, so raw shares in any given minute largely

[3]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

	Share of viewer-minutes			Active, given	Viewer-
	Not tuned in	Passive	Active	tuned in	minutes
During gameplay minute	77.31%	16.77%	5.91%	26.1%	535,171
During commercial minute	77.69%	16.61%	5.71%	25.6%	88,301
During all minutes	77.36%	16.75%	5.88%	26.0%	623,472

Note: Shares of viewer-minutes in each state, for all minutes of the viewer-game pairs in which the viewer was tuned in for at least two minutes; each minute is classified by the state that occupies most of it. "Active, given tuned in" is the active share divided by the tuned-in (passive plus active) share. Commercial minutes are defined at the broadcast level.

reflect where viewers were before the break started. The model in Section 4, which conditions on each viewer's previous choice, shows that commercials substantially reduce both tune-in and active attention.

2.2 Baseball Gameplay Data

We supplement the attention data with pitch-level gameplay data from Data Stadium Inc., which covers all Nippon Professional Baseball (NPB) games in 2018. They record the time stamp and outcome of each pitch and each at-bat, and we use the wall-clock time stamps to merge every at-bat with the minute-level attention data.

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 4.27 runs on average. There are about 314 pitches in each game and each game lasts on average 12,000 seconds.[4] The bottom part of Table 2 reports the same statistics for the 42 broadcast games that we match to the pitch-level data. Among these games, the home team wins with 64.3% probability and scores on average 4.6 runs, while the number of pitches and game length are similar to the full sample. Hence, the home team is more likely to win when the game is broadcast,[5] while other characteristics are similar among games. As the main television networks in Japan are headquartered in Tokyo, the broadcast games tend to feature Tokyo-area teams, such as the Yomiuri Giants and Yakult Swallows. Due to this television presence, Tokyo-area teams tend to have a national fan base.

Figure 2 shows how tune-in and attention evolve over the course of a game. The share of viewers who are tuned in rises steadily from early to late innings. The share of tuned-in viewers who watch actively varies much less across innings and is somewhat higher late in the game. Figure 3 plots the same two measures against the absolute score difference neither showing a monotone pattern. Raw comparisons of this kind mix the effect of the score with the stage of the game, since large margins occur mostly in late innings, and with the composition of games; they motivate the model-based analysis below, which conditions on game and inning.

[4] A game ends up in a tie if both teams have the same score at the end of the 12th inning.

[5] This could happen because those games of top-seed teams when playing home are more likely to be broadcast, so the broadcast sample is selected on this margin.

Table 2: Summary Statistics-Baseball Games

	Mean	SD	Min	Max	N
All games					
Home team win	0.494	–	0	1	877
Home team score	4.273	3.080	0	20	877
Away team score	4.365	3.087	0	20	877
Number of pitches	313.9	39.8	171	495	877
Game length (seconds)	12,002.1	3,062.8	7,623	86,371	877
Broadcast games					
Home team win	0.643	–	0	1	42
Home team score	4.619	3.407	0	16	42
Away team score	3.476	2.549	0	11	42
Number of pitches	306.1	36.8	245	402	42
Game length (seconds)	11,502.6	1,735.5	8,964	16,646	42

Note: A tie is not included in the event of home team winning the game. "Broadcast games" are the televised games in our viewing data that can be matched to the pitch-level data.

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 whichlikely 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, giving our results broader relevance beyond baseball.

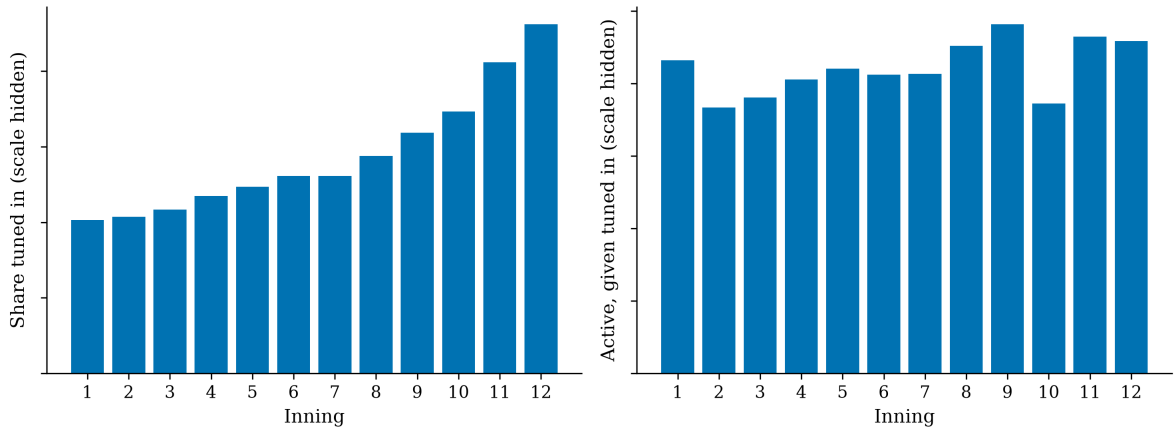
For these reasons, we focus on two key determinants, suspense and surprise, that have been recently formalized by both theoretical and empirical literature. Ely, Frankel, and Kamenica (2015) propose a theoretical model of Bayesian persuasion where the audience derives entertainment utility from suspense 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 situation with bases loaded and two outs is suspenceful, while a surprising moment transpires when a batter hits a home run with no bases loaded.

The construction of suspense and surprise proceeds in two steps, which we describe next.

Step 1: Calculating Winning Probabilities We assume that as they watch the game, viewers form and update beliefs about the binary outcome $\Omega = \{\text{home win}, \text{home does not win}\}$, with a tie counted as the home team not winning.[6] At at-bat t we let $\mu_t \equiv P(hwin|S_t)$ denote the belief that the home team wins (the event "$hwin$"), and we assume viewers who watch the game hold correct beliefs about this probability. μ_t depends on the state variables S_t of the game, which evolve with

[6]One could instead track a richer outcome, such as runs per inning; we follow the existing literature in defining the state space over the final outcome of the game.

Figure 2: Tune-in and Active Attention by Inning



Note: Left: share of viewer-minutes tuned in to the game. Right: share of tuned-in minutes with active attention. Estimation sample. Due to an NDA, we are not allowed to show the scale of the vertical axis; the two panels have different scales.

Markovian probabilities$P(S_{t+1}|S_t)$. At each state we estimate $\hat{P}(hwin|S_t)$ from all 877 games of the 2018 season with XGBoost (Chen and Guestrin, 2016), a standard machine-learning method, tuning its hyperparameters by 5-fold cross validation.[7] Table 3 shows the confusion matrix of our predictive model. Our model correctly classifies the home team's winning with 74.7% accuracy on the 20% random holdout sample, and its precision is 77.8%, which is high compared to win probabilities implied by betting odds of around 60% for experts and users on the leaderboard.[8]

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 (capped at ± 7), and B/O is the pair of bases occupied and out counts. Since there are eight possible configurations of runners on base and three possible out counts, B/O can take 24 different possible values.[9] Although there are potentially other variables that may affect the home team's winning probability such as batting order and star players, we keep the state variables as parsimonious and add other variables in the choice model later.

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

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

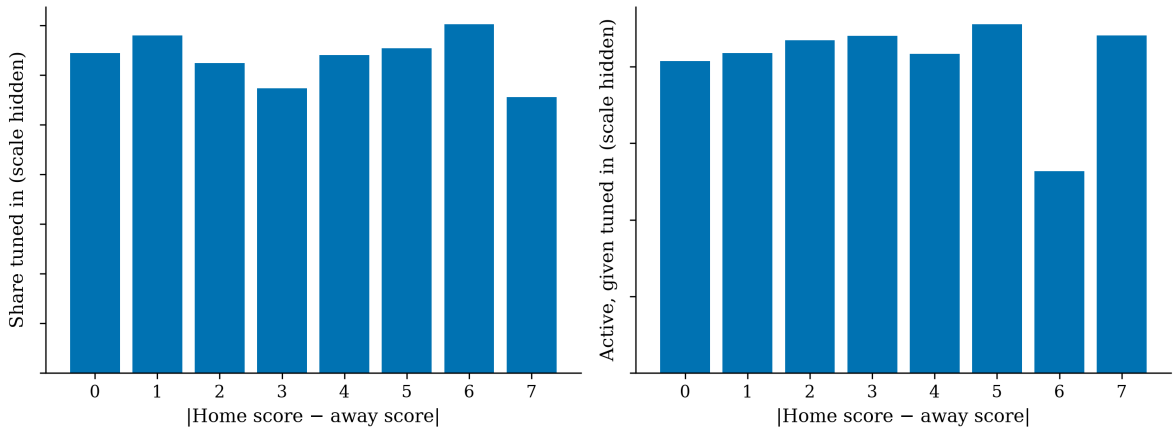
The expectation is calculated with respect to the state transition, $P(S_{t+1}|S_t)$. Suspense is an "ex-ante" measure which is increasing in the variance of the change in the home-win probability between

[7] XGBoost outperformed the random forest, decision tree, and support vector machine alternatives.

[8] For Major League Baseball, see, e.g., https://www.bettingpros.com/mlb/accuracy/game/moneyline/2022/0/. For the Japanese baseball, see, e.g., https://spaia.jp/predictions/npb_ranking/2021.

[9] Our construction of the state variables is based on the large literature of the baseball analytics in Operations Research (e.g., Albert, 2003).

Figure 3: Tune-in and Active Attention by Score Difference



Note: Left: share of viewer-minutes tuned in to the game. Right: share of tuned-in minutes with active attention, by the absolute score difference (capped at 7). Estimation sample. Due to an NDA, we are not allowed to show the scale of the vertical axis; the two panels have different scales.

Table 3: Confusion Matrix

Model / Data	0 (home does not win)	1 (home win)
0 (home does not win)	23,653	8,920
	(43.0%)	(16.2%)
1 (home win)	5,000	17,486
	(9.1%)	(31.8%)

The table reports the confusion matrix on a 20% random holdout sample of at-bats. Rows show the classification based on the model prediction and columns the observed outcome. The numbers in parentheses are the fraction of each case.

the next and current at-bats. In contrast, surprise is "ex-post" and generated by unanticipated events, defined as the absolute distance between the current and previous beliefs:[10]

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

To compute suspense we need the state-transition probability $P(S_{t+1}|S_t)$, which we estimate directly from the play-by-play data. The inning and score-difference components of the transition are deterministic given the change in the base-out configuration, so that only the 24×24 transition matrix over base-out states must be estimated, separately for each inning, half-inning, and score difference. We defer the details of this construction to Appendix A. Surprise requires no transition probabilities, being simply $SURP_t = |\hat{\mu}_t - \hat{\mu}_{t-1}|$.

Summary Statistics of Suspense and Surprise Table 4 reports summary statistics of the suspense and surprise measures, computed from all 877 games in 2018 regardless of whether a

[10] We follow the specification used in Bizzozero, Flepp, and Franck 2016.

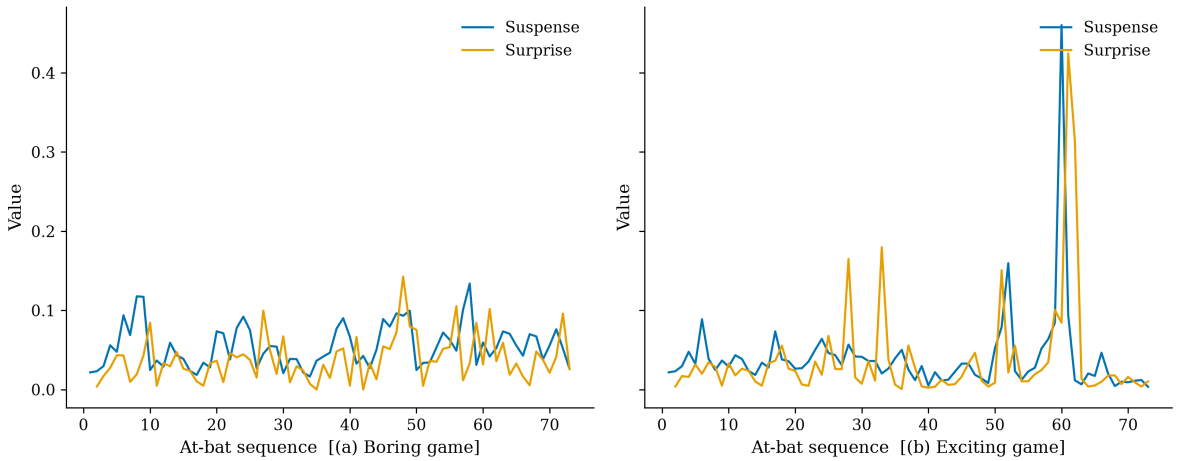
Table 4: Summary Statistics: Suspense and Surprise

	Suspense	Surprise
Mean	0.040	0.030
Std. dev.	0.037	0.043
Median	0.031	0.018
Obs.	68,881	68,234

Note: The suspense and surprise measures are computed at the at-bat level with all baseball games in 2018.

game was broadcast. The average suspense is 0.040 and the average surprise is 0.030, and the two measures have similar standard deviations (0.037 and 0.043).[11] The distribution of surprise is more skewed: its median, 0.018, is well below its mean, because most at-bats move the win probability only slightly while a few, such as home runs in close games, move it a lot.

Figure 4: Examples of Suspense and Surprise Over Time



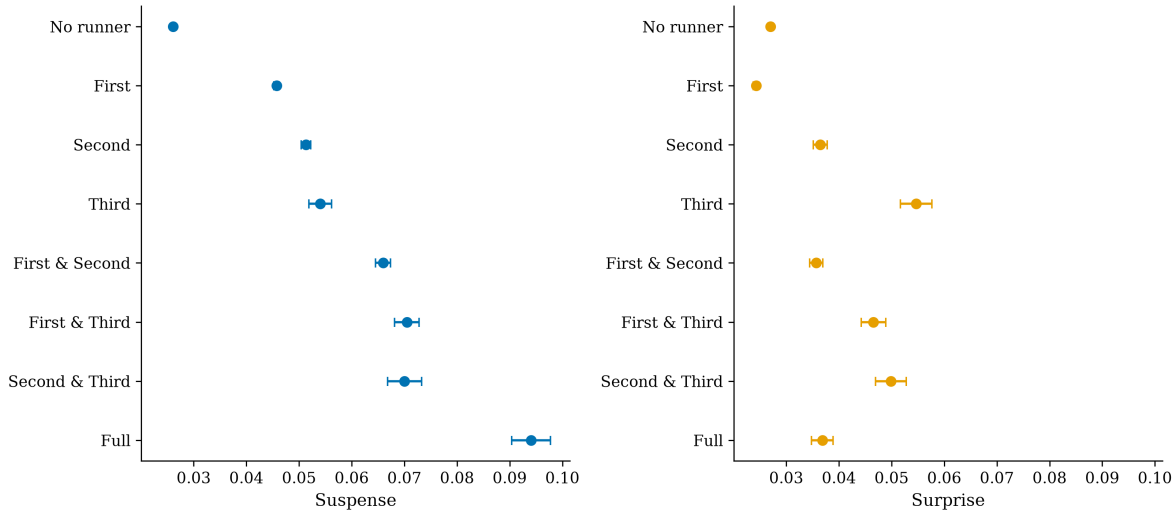
Note: Suspense and surprise by at-bat in two games: (a) a game decided early, (b) a close game decided late.

Figure 4 shows how the two measures move within two exemplar games. In panel (a), one team takes control early and the win probability moves little, so suspense and surprise remain low and fluctuate within a narrow band throughout. By contrast, the game in panel (b) remais close up to the end. The win probability swings sharply in the late innings, producing the largest values of both suspense and surprise then.

Figure 5 highlights some key differences in what is measured by suspense vs. surprise. The left-hand panel shows that suspense rises with the number of runners on base and is highest 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, the right-hand panel shows that surprise, which

[11] Surprise is not defined for the first at-bat of a game. Suspense is not defined for the 230 at-bats (0.3%) whose state is never followed by another at-bat in the transition data, most of them the final at-bat of a game.

Figure 5: Suspense and Surprise by Runners on Base



Note: The figures plot the average and the 95% confidence intervals for suspense (left) and surprise (right) by the configuration of runners on base.

is an "ex-post" measure, does not rise systematically with the runners on base: it is highest with a runner on third and, for example, no higher with the bases loaded than with a runner on second. Additionally, suspense and surprise comove with other game features, such as the inning and the score difference; we report these patterns in Appendix A (Figures A.1 and A.2).

4 Viewers' Tune-in and Attention Choice Model

We model each viewer as making two linked choices in every minute t of a game she watches: whether to be tuned in to the game, $T_{it} \in \{0,1\}$, and, if tuned in, whether to watch actively, $A_{it} \in \{0,1\}$. Expected content enters the tune-in decision and realized content enters the attention decision, reflecting the assumption that the game state is revealed only to a viewer who is actively watching.[12] To simplify notation we useS_t to denote the realized suspense SUS_t and $|U_t|$ for the realized surprise $SURP_t$, with $\bar{S}_{it}$ and $\bar{U}_{it}$ for their expected counterparts defined below.

4.1 Viewers' Information and Beliefs

Viewers' decisions to tune in and exert active attention are made with different information. A viewer who is tuned in sees the game on the screen, so her attention decision can respond to the current game content. A viewer who is not tuned in, or who is tuned in but looking away, does not see the current play. We assume that the game state is revealed to a viewer only in minutes in which she actively watches game content (active minutes during a commercial break reveal nothing about the game). Let τ_{it} denote the last minute before t in which viewer i actively watched game

[12]Section H in the Appendix examines the sensitivity of the estimates to this information assumption, including a specification that grants a viewer who was active in the previous minute advance information about the current play.

content, and s_τ the game state she saw then. Between τ_{it} and t, the game advances by a random number of at-bats. We describe her belief about the current state by

$$\pi_{it} \;=\; \sum_{k\geq 0} \Pr\left(k \mid t-\tau_{it}\right)\; e(s_\tau)' P^k, \tag{3}$$

where $e(s_\tau)$ is the indicator vector of the state she last saw, P is the at-bat transition matrix of Section 3 augmented with the probability that the game ends, and $\Pr(k \mid g)$ is the probability that k at-bats occur in g minutes, which we estimate from the at-bat time stamps of all 877 games. On average 0.39 at-bats occur per minute, and no at-bat occurs in 61 percent of minutes. Our preferred specification propagates the last observed state over every subsequent gap in the game. For viewers who have not yet watched in the current game, we instead use the empirical distribution of game states at the same point of play across all 877 games.

From these beliefs we compute the content the viewer *expects*. Expected suspense is $\bar{S}_{it} = \sum_s \pi_{it}(s)\, SUS(s)$, and expected surprise $\bar{U}_{it}$ is the expected absolute change in the win probability in the most recent at-bat the viewer has not seen (zero if no at-bat is expected to have occurred). The content a tuned-in viewer *sees* is the realized suspense S_t and realized surprise $|U_t|$ of the current at-bat. During commercial breaks the game is not on screen, and we hold realized content at its value just before the break.

4.2 Choice Probabilities

We model the probability of tuning in is a logit that depends on expected content,

$$\begin{aligned}
\Pr(T_{it}=1) &= \Lambda\left(V^T_{it}\right), \\
V^T_{it} &= \beta^T_S \bar{S}_{it} + \beta^T_U \bar{U}_{it} + \rho\, T_{i,t-1} + \gamma^T_0 CM_t + \gamma^T_1 CM_{t-1} \\
&\quad + x_i'\theta^T + \phi^T_H H_{ig} + \phi^T_R R_{ig} + \delta^T_{it},
\end{aligned} \tag{4}$$

where Λ is the logistic (soft-max) function, CM_t indicates a commercial minute, x_i contains age and gender, H_{ig} and R_{ig} indicate that viewer i is a fan of the home or the away team in game g, and δ^T_{it} collects fixed effects described below. The lagged choice $T_{i,t-1}$ captures the strong persistence of viewing. Conditional on tuning in, the probability of watching actively depends on the content on screen, with slopes that differ by fandom,

$$\begin{aligned}
\Pr(A_{it}=1 \mid T_{it}=1) &= \Lambda\left(V^A_{it}\right), \\
V^A_{it} &= \left(\beta_S + \beta^H_S H_{ig} + \beta^R_S R_{ig}\right) S_t + \left(\beta_U + \beta^H_U H_{ig} + \beta^R_U R_{ig}\right) |U_t| \\
&\quad + \lambda_A A_{i,t-1} + \lambda_P P_{i,t-1} + \gamma^A_0 CM_t + \gamma^A_1 CM_{t-1} + x_i'\theta^A + \phi^A_H H_{ig} + \phi^A_R R_{ig} + \delta^A_{it},
\end{aligned} \tag{5}$$

where $A_{i,t-1}$ and $P_{i,t-1}$ indicate active and passive attention in the previous minute. Neutral viewers are the reference group, so that β_S and β_U are their slopes, and the slopes of home-team

and away-team fans are the sums $\beta_S + \beta_S^H$, $\beta_S + \beta_S^R$, and analogously for surprise.

Both indices include fixed effects for the game, the clock hour of the broadcast, the inning, and the time (in minutes) since the viewer last watched actively. The game fixed effects absorb differences across games in team quality, fan base, and broadcast slot; the inning fixed effects absorb the systematic evolution of the audience over a game (Figure 2); and the information-recency fixed effects absorb how long a viewer has been away from the screen.

The probabilities of the three outcomes are $\Pr(T_{it} = 0) = 1 - \Lambda(V_{it}^T)$, $\Pr(\text{passive}) = \Lambda(V_{it}^T)\,[1 - \Lambda(V_{it}^A)]$, and $\Pr(\text{active}) = \Lambda(V_{it}^T)\,\Lambda(V_{it}^A)$. Because the two indices share no parameters, the likelihood factors into a logit for tune-in, estimated on all viewer-minutes, and a logit for active attention, estimated on tuned-in minutes; we estimate each by maximum likelihood and cluster standard errors by game. Beliefs are always computed on each viewer's complete timeline, including minutes that do not enter a given likelihood.[13]

The model is myopic, as viewers respond to the content in front of them but do not weigh how watching now changes what they will know later. Dynamics still enter through the lagged choices and through beliefs, which depend on the viewer's own history of attention, and this is what lets us simulate counterfactual commercial schedules, since a schedule changes who watches when, which changes what each viewer knows about the game and hence her later choices. In Appendix app:dynamic we specify and estimate a fully forward-looking models where viewers tune in to gain information about the likely future path of the game. We find that it does not predict the data better than the myopic model presented here.

4.3 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. We include game fixed effects to control for differences across games, such as the level of skill or fandom 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. We further include inning fixed effects to control for particular stages of the game. The remaining variation in suspense and surprise is due to the particular (random) score path realized within a game. More formally, this setting leads to standard exclusion restrictions that allow us to identify viewers' tastes for suspense and surprise:

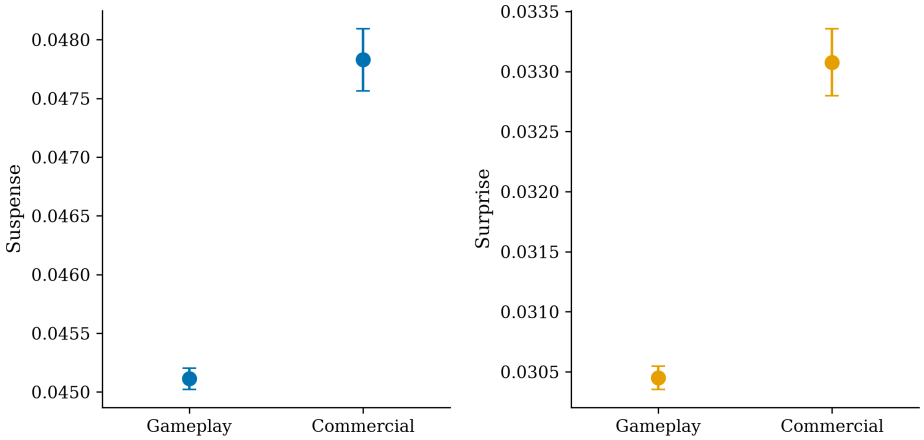
$$E(\varepsilon_{it}^A\, S_t \mid CM_t, x_{it}, \delta_{it}^A) = 0, \qquad E(\varepsilon_{it}^A\, |U_t| \mid CM_t, x_{it}, \delta_{it}^A) = 0,$$

[13] An earlier version of this paper estimated a nested logit in which the passive and active alternatives form a nest. In our data the dissimilarity parameter of that model is unstable across specifications and falls outside the unit interval required for consistency with random utility maximization. The sequential specification above is the special case in which the tune-in and attention errors are independent, which avoids estimating that parameter at the cost of ruling out a common unobservable across the two margins.

and analogously for expected suspense and surprise in the tune-in equation.

Commercials The timing of commercials is largely predetermined: in Japanese baseball telecasts, breaks are aired at half-inning changes and other stoppages, and advertisers buy slots at the game level rather than at particular game situations. Figure 6 compares suspense and surprise in commercial and gameplay minutes. The content carried into commercial minutes is slightly higher, by about 6 percent for suspense and 9 percent for surprise, differences that are small relative to the within-game variation of the two measures (standard deviations of about 0.044). The game and inning fixed effects absorb the systematic placement of breaks over the course of a game.

Figure 6: Suspense and Surprise in Gameplay and Commercial Minutes



Note: Averages and 95% confidence intervals of minute-level suspense (left) and surprise (right) in gameplay and commercial minutes. During commercial minutes, the measures are held at their value just before the break. The vertical axes do not start at zero.

4.4 Results

Table 5 reports the estimates. The estimation sample has 380,092 viewer-minutes for the tune-in equation and 48,298 tuned-in minutes for the attention equation; 43 games contribute to the attention likelihood.

Content. Expected suspense raises the probability of tuning in (coefficient 2.06; $p = 0.063$) while expected surprise has no significant impact. Among tuned-in viewers, realized surprise raises the probability of active attention (0.92, $p < 0.001$ for neutral viewers), while realized suspense has a positive but insignificant effect for neutral viewers. Suspense and surprise therefore act on different margins: suspense on the extensive margin (tuning-in) and surprise on the intensive margin (active attention).

Fandom. The response to realized content differs sharply by fandom (Table 6). For home-team fans, the suspense impact is about six times that of neutral viewers (3.21, $p = 0.014$): a

Table 5: Estimates of the Tune-in and Attention Model

	(1) Tune in	(2) Active | tuned in
Expected suspense $\bar{S}_{it}$	2.060^{*}	
	(1.110)	
Expected surprise $\bar{U}_{it}$	0.576	
	(1.620)	
Realized suspense S_t		0.553
		(0.438)
Realized surprise $\lvert U_t \rvert$		0.915^{***}
		(0.223)
Commercial CM_t	-0.463^{***}	-0.347^{***}
	(0.041)	(0.061)
Commercial CM_{t-1}	0.083^{*}	-0.119^{**}
	(0.045)	(0.054)
Tuned in at $t-1$	5.558^{***}	
	(0.051)	
Active at $t-1$		1.409^{***}
		(0.121)
Passive at $t-1$		-1.485^{***}
		(0.050)
Age	0.003^{***}	0.003^{**}
	(0.001)	(0.001)
Female	-0.068^{***}	0.045
	(0.026)	(0.042)
Home fan H_i	0.052	-0.090
	(0.046)	(0.104)
Away fan R_i	0.039	0.157
	(0.051)	(0.199)
$S_t \times H_i$		2.654^{**}
		(1.353)
$\lvert U_t \rvert \times H_i$		-0.343
		(0.721)
$S_t \times R_i$		1.367
		(3.363)
$\lvert U_t \rvert \times R_i$		-8.329^{***}
		(3.033)
Fixed effects	Game, clock hour, inning, information recency	
Observations	380,092	48,298
Log likelihood	$-51{,}342.8$	$-16{,}749.3$

Note: Logit estimates of equations (4) and (5). Column (1): all viewer-minutes; column (2): tuned-in minutes. Standard errors clustered by game in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 6: Content Effects on Active Attention by Viewer Group

	Realized suspense		Realized surprise		Tuned
Viewer group	Slope	+1 SD (pp)	Slope	+1 SD (pp)	minutes
Neutral	0.553	+0.25	0.915***	+0.42	39,201
	(0.438)		(0.223)		
Home fan	3.207**	+1.59	0.572	+0.28	7,293
	(1.302)		(0.686)		
Away fan	1.919	+0.84	−7.414**	−3.04	1,804
	(3.371)		(2.996)		

Note: "Slope" is the coefficient on realized suspense or surprise for each group implied by column (2) of Table 5 (neutral: main effect; home and away fans: main effect plus interaction), with standard errors clustered by game. "+1 SD (pp)" is the average change in the predicted probability of active attention, in percentage points, when the content measure rises by one sample standard deviation (0.044 for both measures), holding all other covariates fixed. The last column reports the number of tuned-in minutes in each group.

one-standard-deviation increase in suspense raises their probability of active attention by 1.59 percentage points, compared with 0.25 points for neutral viewers. Away-team fans respond in the opposite direction to surprise (slope -7.41, $p = 0.013$): a one-standard-deviation increase in surprise lowers their probability of active attention by 3.04 percentage points, while it raises that of neutral viewers by 0.42 points. Because surprise is measured in absolute value, it combines good and bad news for any given team; the negative impact for away-team fans shows that, on average, large swings reduce their attention. The four fandom interactions are jointly significant on a conventional Wald test ($\chi^2(4) = 16.81$, $p = 0.002$), though with only 43 game clusters this test is optimistic; a restricted-score multiplier bootstrap gives $p = 0.045$–0.047, and we discuss few-cluster inference in Section H.

Persistence and demographics. Both choices are highly persistent. Being tuned in in the previous minute raises the log odds of being tuned in by 5.56, and active attention in the previous minute raises the log odds of active attention by 1.41 relative to a viewer who was not tuned in, while passive attention lowers it by 1.49. Older viewers are more likely to tune in and to watch actively, and women are somewhat less likely to tune in. To assess whether this persistence reflects genuine state dependence rather than unobserved viewer heterogeneity, we estimated a model including a latent-type mixture with two or three viewer types to control for viewer heterogeneity (see Heckman (1991)). The results, in (Appendix F), show that the persistence coefficients change little, suggesting that the estimated persistence represents intrinsic state dependence.

Commercials. Commercials reduce both margins. A commercial minute lowers the odds of being tuned in by 37 percent ($\exp(-0.463) - 1$) and, for a tuned-in viewer, lowers the odds of active attention by 29 percent ($\exp(-0.347) - 1$); a commercial in the previous minute further lowers active attention. Because these effects are conditional on the previous minute's choice, they describe how commercials change transitions: commercials reduce both tune-in and active attention

and, given the state dependence in viewers' choices, have persistent negative impacts on viewership and attention even after the commercial break.

5 Counterfactual Exercises: The Timing of Commercial Breaks

We now use the estimated model to study a design lever that broadcasters control directly: when commercial breaks are aired. The estimates imply that the timing of breaks matters for two reasons. First, commercials reduce tune-in and attention, and because both are persistent, a break placed at a moment when many viewers are actively watching loses more attention than one placed when few are. Second, viewers who look away during a break lose track of the game, so a break changes what viewers know and therefore their later choices. The broadcaster faces a trade-off: placing breaks when viewers are attentive raises attention to the commercials but interrupts the game when it is most engaging.

5.1 Design of the Counterfactuals

Feasible schedules. We hold fixed each game's observed commercial blocks: their number, their lengths, and therefore the total commercial time (676 commercial minutes in the 43 games in which the attention equation is identified). The broadcaster chooses the start minute of each block among the observed start minutes and the minutes around each half-inning change, which are the natural stoppages in play. Blocks may not overlap, and consecutive blocks must be separated by at least one minute of play.

Simulation. For a given schedule, we simulate every viewer minute by minute from the start of each game. In each minute, tune-in and attention are drawn from equations (4) and (5). A simulated active minute during game play reveals the game state and updates the viewer's beliefs and information recency; the simulated choices become next minute's lagged choices. The game itself is unaffected by the broadcast schedule, but the content on screen is not: during a commercial break the screen does not show the game, so realized content is held at its pre-break value at each schedule's own breaks. We calibrate the two intercepts so that the simulation under the observed schedule reproduces the observed numbers of tuned-in and active minutes, and evaluate every schedule with the same 500 sets of random draws. We report expected active viewer-minutes in total, during commercials, and during game play ("attention to the game"). Under the observed schedule these are about 10,170, 1,160, and 9,010 viewer-minutes.

Policies. We consider four kinds of schedules. (i) *Total-attention optimum.* We first solve, game by game, the integer program that maximizes expected active attention one minute ahead given observed histories (the "one-step" optimum), and then improve it by a local search over block relocations that evaluates each candidate with the full recursive simulation (the "recursive" optimum). (ii) *Audience/advertising frontier.* For each game we build a library of candidate

Table 7: Optimizing Break Times for Total Active Attention

	Active viewer-minutes			Active commercial viewer-minutes	
Comparison	Change	%	95% CI (%)	Change	95% CI
One-step optimum vs. observed	+168.8	+1.66%	[0.83, 2.66]	−264.4	[−393.0, −152.3]
Recursive optimum vs. observed	+213.0	+2.09%	[1.26, 3.09]	−237.9	[−364.0, −128.9]
Recursive vs. one-step optimum	+44.3	+0.43%	[0.23, 0.68]	+26.5	[7.3, 47.9]

Note: Changes in expected active viewer-minutes, summed over the 43 games, relative to the observed schedule (rows 1–2) or to the one-step optimum (row 3). Confidence intervals from 9,999 game-bootstrap replications, holding coefficients and schedules fixed. 500 simulation draws.

schedules from local searches that weight total attention and attention to commercials differently, and then choose one schedule per game to maximize total active attention subject to a minimum level of active attention during commercials, relative to the observed schedule. (iii) *Ad-attention optimum.* The schedule that maximizes active attention during commercials, with and without a constraint that attention to the game may not fall. (iv) *Simple rules.* Schedules that place breaks to minimize the realized suspense (or surprise) shown during commercial minutes, a variant that only allows each block to move by up to 15 minutes, and a rule that places breaks in innings 7 and later wherever possible. The optimization is a local search, so the optimized schedules are lower bounds on what an unconstrained optimum could achieve within our design.

Inference. Our main confidence intervals resample the 43 games (9,999 game-bootstrap replications) holding the estimated coefficients and schedules fixed. To account for estimation uncertainty, we also draw 50 game-cluster bootstrap samples, re-estimate every coefficient and fixed effect of the model in each, and re-evaluate the main policies.

5.2 Results

Total attention. Table 7 reports the effect of optimizing break times for total active attention. The one-step optimum raises active attention by 1.66 percent, and the recursive optimum by 2.09 percent (95% confidence interval 1.26 to 3.09 percent), or about 213 active viewer-minutes over the 43 games, without any change in advertising time. Accounting for how a schedule changes later choices and beliefs adds about a quarter to the one-step gain (0.43 percentage points, interval 0.23 to 0.68). These gains come at the expense of attention to commercials: the recursive optimum lowers active attention during commercials by about 238 viewer-minutes, about a fifth of the observed level, because it places breaks where few viewers would be actively watching anyway.

The audience/advertising frontier. The loss of attention to commercials is not necessary for most of the gain. Table 8 and Figure 7 trace the frontier between total active attention and attention to commercials. The schedule that maximizes total active attention without any constraint raises it by 2.84 percent while retaining 88 percent of active attention during commercials. Requiring that attention to commercials not fall below its observed level still allows an increase in total active

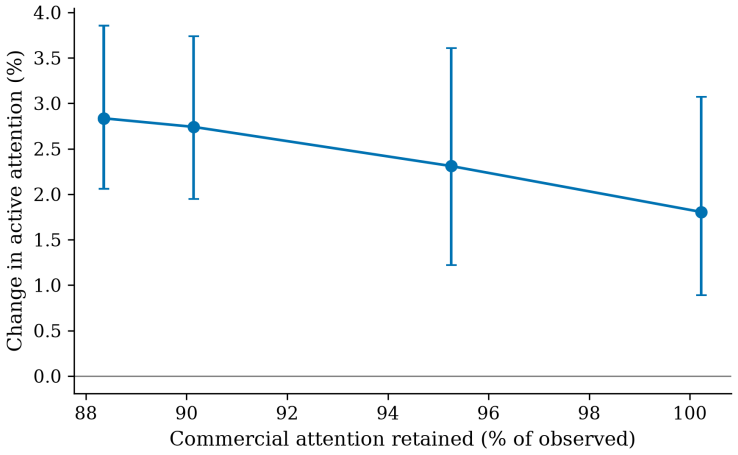
Table 8: The Audience/Advertising Frontier

	Change in active attention		Commercial attention retained	
Schedule	%	95% CI	Share	95% CI
Unconstrained maximum	+2.84%	[2.06, 3.86]	88.4%	[79.5, 95.6]
Minimum 80% ad attention	+2.84%	[2.06, 3.86]	88.4%	[79.5, 95.6]
Minimum 85% ad attention	+2.84%	[2.06, 3.86]	88.4%	[79.5, 95.6]
Minimum 90% ad attention	+2.74%	[1.95, 3.74]	90.1%	[81.5, 97.7]
Minimum 95% ad attention	+2.31%	[1.22, 3.61]	95.3%	[83.6, 105.5]
Minimum 100% ad attention	+1.80%	[0.89, 3.07]	100.2%	[90.0, 108.1]

Note: Each row maximizes total active attention subject to retaining at least the stated share of the observed active attention during commercials. "Commercial attention retained" is the resulting active attention during commercials as a share of the observed level. Confidence intervals from the game bootstrap. The constraints at 80% and 85% do not bind.

attention of 1.80 percent (interval 0.89 to 3.07). For a broadcaster, retiming breaks is therefore close to a free improvement in audience engagement: it raises attention to the program without reducing, on average, the attention that advertisers buy.

Figure 7: The Audience/Advertising Frontier

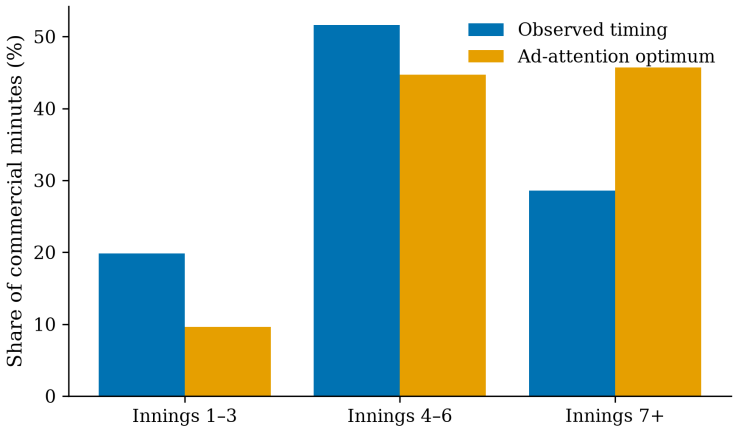


Note: Change in total active attention against active attention during commercials retained, for the schedules in Table 8. Bars are game-bootstrap 95% confidence intervals.

Maximizing attention to commercials. Table 9 reports the remaining policies. Choosing break times to maximize active attention during commercials raises it by 7.84 percent (about 91 viewer-minutes, interval 21 to 190). The cost to the game is small and uncertain: attention to the game falls by 0.75 percent, with an interval that includes zero (−212 to +53 viewer-minutes), and each additional minute of attention to commercials displaces about 0.75 minutes of attention to the game. The broadcaster can also obtain most of this gain with no loss of attention to the game at all: the best schedule that does not reduce attention to the game raises attention to commercials by 86 viewer-minutes (7.4 percent).

Figure 8 shows how the ad-attention schedule achieves its gain. It moves 445 of the 676 commercial minutes, shifting breaks out of the early innings and into the late innings: the share of commercial minutes in innings 7 and later rises from 29 to 46 percent. Evidently, to maximize ad-attention, we would optimally place breaks when more viewers are tuned in and actively watching, which is late in the game.

Figure 8: Placement of Commercial Minutes by Inning



Note: Share of commercial minutes in each group of innings under the observed schedule and under the schedule that maximizes active attention during commercials.

Simple rules. The simple rules in Table 9 show which features of the optimal schedules matter. Placing breaks in low-suspense stretches of the game raises attention to the game by about 311 viewer-minutes (3.5 percent; interval 163 to 500) and total active attention by 2.1 percent, at the cost of about 100 viewer-minutes of attention to commercials. The gain is nearly as large when each block may move by at most 15 minutes. A rule based on surprise does somewhat less well, and a rule that moves breaks into the late innings raises attention to the game without a significant loss of attention to commercials. These rules are easy to implement in a live broadcast, since suspense is computable in real time from the game state.

Parameter uncertainty. Table 10 re-evaluates the main policies in 50 bootstrap replicates that re-estimate every coefficient and fixed effect of the model. The recursive optimum raises active attention in every replicate (2.07 percent, 95 percent interval 1.04 to 3.74), as do both frontier schedules and the low-suspense rule; the best ad-attention schedule that protects attention to the game does so in 98 percent of replicates. The ad-attention optimum raises attention to commercials in every replicate, but raises total active attention in only two thirds of them, consistent with its small and uncertain effect on attention to the game.

Table 9: Attention to Commercials and the Game under Alternative Schedules

	Total active		Active during commercials		Active to the game	
Schedule	Change	95% CI	Change	95% CI	Change	95% CI
Ad-attention optimum	+22.9	[−45.4, 102.1]	+90.9	[20.9, 189.8]	−68.0	[−211.6, 53.3]
Ad optimum, no loss of game attention	+90.5	[31.6, 162.6]	+85.5	[17.7, 183.4]	+5.0	[−124.4, 117.1]
Low-suspense rule	+210.8	[112.1, 328.7]	−100.4	[−182.7, −33.0]	+311.2	[162.6, 500.5]
Local low-suspense rule	+192.5	[89.3, 325.3]	−119.1	[−213.4, −48.5]	+311.6	[144.4, 532.9]
Low-surprise rule	+140.1	[61.4, 230.2]	−95.0	[−154.1, −44.4]	+235.0	[119.7, 369.3]
Late-inning placement	+166.3	[78.7, 268.9]	−27.1	[−68.2, 19.1]	+193.4	[77.2, 320.2]

Note: Changes in expected active viewer-minutes, summed over the 43 games, relative to the observed schedule. "Ad optimum, no loss of game attention" maximizes active attention during commercials subject to no reduction in active attention to the game. Confidence intervals from the game bootstrap. 500 simulation draws.

Table 10: Re-estimation Bootstrap

	Active attention (%)		Commercial attention (viewer-min.)		Share of replicates with a gain	
Policy	Point	95% interval	Point	95% interval	Active	Commercial
Recursive optimum	+2.07%	[1.04, 3.74]	−235	[−349, −115]	1.00	0.00
Pareto: unconstrained maximum	+2.97%	[1.86, 4.27]	−130	[−222, −35]	1.00	0.00
Pareto: ad attention held at 100%	+1.75%	[0.87, 2.89]	+10	[−74, 111]	1.00	0.50
Ad-attention optimum	+0.30%	[−0.72, 0.91]	+95	[32, 206]	0.66	1.00
Low-suspense rule	+1.84%	[1.32, 2.79]	−100	[−147, −40]	1.00	0.00
Protect baseball: max loss 0%	+1.02%	[0.20, 1.49]	+90	[5, 198]	0.98	0.98

Note: 50 game-cluster bootstrap samples; in each, every coefficient and fixed effect of the model is re-estimated and the policies are re-evaluated. Schedules are held fixed at those chosen with the full-sample estimates. Point estimates use the full-sample estimates.

Table 11: Counterfactual: Shortening the Game to Seven Innings

	Active attention per minute			Total active
Viewer group	Observed	7 innings	% change	minutes, %
Gameplay minutes				
All	2.81%	2.74%	−2.5%	−20.1%
Male	2.87%	2.79%	−2.7%	−20.6%
Female	2.72%	2.66%	−2.1%	−19.5%
Commercial minutes				
All	1.98%	1.96%	−1.1%	−19.0%
Male	2.00%	1.97%	−1.6%	−19.6%
Female	1.96%	1.95%	−0.3%	−18.2%

Note: Active attention per minute, and the change in total active viewer-minutes, under a seven-inning game (innings three through nine of each observed game, relabeled one through seven) relative to the observed nine-inning game, in the 43 attention-identified games. Suspense and surprise are recomputed under a seven-inning win-probability surface; each viewer's history is reset at the new first inning. "% change" is the change in active attention per minute; "Total active minutes, %" is the change in the summed active viewer-minutes.

5.3 Shortening the Game

A second design lever, discussed by professional leagues for the pandemic-shortened 2020 season, is to shorten the game from nine innings to seven. Its effect on attention has two margins that pull in opposite directions. On the *extensive* margin, a shorter broadcast has fewer minutes in which to accumulate attention, which lowers the total. On the *intensive* margin, compressing the game is thought to concentrate the drama: with fewer innings left, each remaining play is more decisive, so suspense per minute should be higher and viewers should attend more closely while they watch. The case for shortening as a way to raise engagement rests on the intensive-margin gain outweighing the extensive-margin loss.

Our estimates run clearly against that hypothesis: the intensive-margin gain is close to absent, so there is little to offset the mechanical extensive-margin loss. Table 11 reports the counterfactual, which retains each game's innings three through nine and relabels them one through seven (a seven-inning game treated as a nine-inning game resumed in the third); the construction of the seven-inning win-probability surface and a decomposition of the result into its channels are in Appendix E. The extensive margin is large and negative, as expected: total active attention falls by about 20 percent, for both gameplay and commercial minutes, close to the mechanical consequence of dropping about 18 percent of the minutes. The intensive margin, however, manifests minimally, as active attention *per minute* falls slightly, by about 2.5% during gameplay and 1% during commercials. What limits the intensive margin is that shortening the game has negligible impacts on the "drama" of the game, as mean per-at-bat suspense rises only from 0.033 to 0.035. This is not enough to drive the incremental attention needed to counteract the large negative extensive margin impact.

5.4 Robustness to the Information Assumption

Our main specification assumes a viewer learns the game state only by actively watching. Appendix H assesses alternatives where 1) passive presence also reveals the state, and 2) a viewer who was active in the previous minute is granted the current play's realized suspense and surprise. The content estimates are qualitatively stable across these assumptions: expected suspense raises tune-in and realized surprise raises active attention throughout. The appendix also reports sensitivity to the belief-memory rule, few-cluster (wild and score-multiplier) inference for the fandom interactions, which weakens their conventional significance, and recursive re-evaluations of the commercial-timing policies under each information assumption, whose qualitative conclusions are unchanged.

6 Conclusion

In this paper, we show that distinguishing active from passive attention changes what standard viewing measures reveal about television audience engagement. Using a unique dataset that tracks viewers' faces at high frequency, we separate, minute by minute, whether a viewer is not tuned in, passively present, or actively watching. Tuned-in viewers actively watch only about one minute in five. We model tuning in and attention as two linked choices made with different information, since the game state is revealed only to a viewer who is actively watching. Expected suspense raises the probability of tuning in, while realized surprise raises the probability that a tuned-in viewer watches actively; suspense and surprise therefore act on different margins of engagement, and the response to realized content varies by fandom. Commercials lower both tune-in and active attention. Using the estimated model to retime commercial breaks, we find that choosing break times to maximize total active attention raises it by about 2.1 percent, most of it with no loss of attention to the commercials, while a schedule that maximizes attention to the commercials raises it by about 7.8 percent with an uncertain change in attention to the game.

Our results provide managerial implications to different stakeholders. For broadcasters, the timing of breaks is a low-cost lever: simple rules that place breaks in low-suspense stretches of play recover much of the attainable gain in attention to the game. For advertisers, our results suggest that slots late in the game, when more viewers are tuned in and attentive, deliver more attention per commercial minute, and that attention rather than viewing should guide the valuation of slots. More broadly, our study highlights the value of granular data on consumer attention, without which these managerial insights cannot be found.

Several issues are left for future research. First, this analysis could be extended to other genres of TV programming beyond baseball. 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, future research could examine content factors beyond suspense and surprise, such as auditory and/or visual salience. Lastly, future research could link the price of commercials to active versus passive attention and examine broadcasters' joint choice of break timing and pricing.

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. 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.

Heckman, J. J. (1991): "Identifying the hand of past: Distinguishing state dependence from heterogeneity," *The American Economic Review*, 81(2), 75–79.

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.

Tuchman, A. E., H. S. Nair, and P. M. Gardete (2018): "Television ad-skipping, consumption complementarities and the consumer demand for advertising," *Quantitative Marketing and Economics*, 16(2), 111–174.

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.

Yao, S., W. Wang, and Y. Chen (2017): "TV Channel Search and Commercial Breaks," *Journal of Marketing Research*, 54(5), 671–686.

Appendix A Construction of Suspense, Surprise, and Beliefs

Transition probabilities. We estimate the at-bat transition matrix from consecutive at-bats in all 877 games of the 2018 season, separately for each inning, half-inning, and score difference (capped at ± 7), over the 24 base-out configurations. Given a transition between base-out configurations, the rest of the next state is deterministic. A half-inning ends whenever the number of outs falls, which covers both the third out of an inning and inning-ending double plays; no runs are scored on such a transition, the batting team changes, and after the bottom half the inning advances. Otherwise, the number of runs scored equals one (the batter) plus the change in the number of runners and outs, and the score difference moves in favor of the batting team.

Suspense. For each at-bat, suspense is the square root of the expected squared change in the home-win probability over the next at-bat, taking the expectation over the estimated transitions and evaluating the home-win probability at each possible next state with the XGBoost model of Section 3. Transitions out of a game's final at-bat are never observed, so the 230 at-bats whose state is never followed by another at-bat in the data (0.3 percent of at-bats, most of them final at-bats) have no defined suspense and are excluded from the analysis.

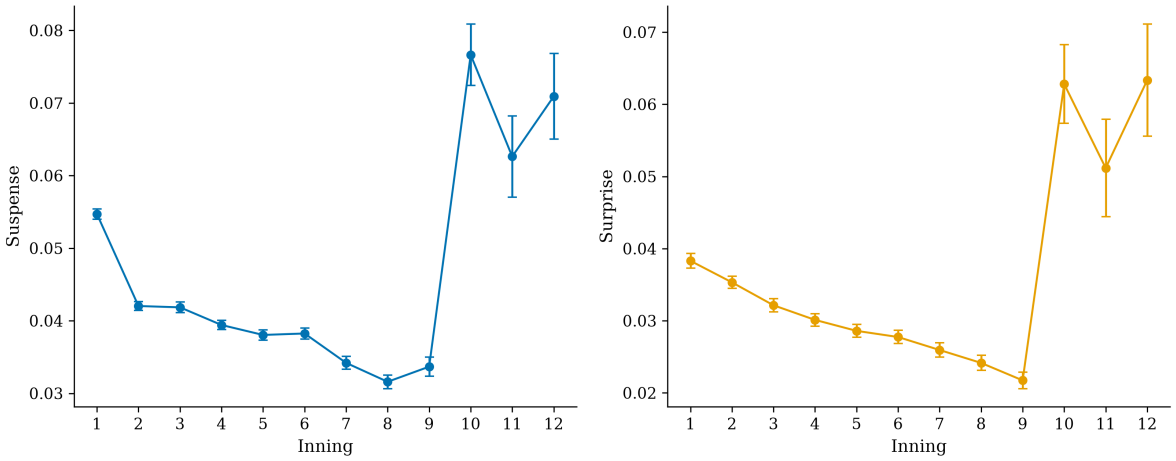
From at-bats to minutes. We place each at-bat in a minute of the broadcast using the time stamp of its first pitch, assigning pitches after the 30th second to the next minute. When several at-bats start in the same minute, the minute takes the state of the last of them; minutes without a new at-bat carry the previous at-bat's state. During commercial minutes, the realized suspense and surprise are held at their values just before the break.

Beliefs. The belief in equation (3) propagates the last state a viewer saw by the at-bat transition matrix, augmented with the probability that a state is a game's last at-bat, which we estimate from the final at-bats of the 877 games. The distribution of the number of at-bats k in a gap of g minutes, $\Pr(k \mid g)$, is the empirical distribution from the at-bat time stamps of the same games; the expected number of at-bats is 0.39 for a one-minute gap and 3.9 for a ten-minute gap. We validated the exact belief moments against an independent Monte Carlo simulation of the game from 500 anchor states: all 500 simulated expectations of suspense lie within four Monte Carlo standard errors of the exact values (correlation 0.99998).

Appendix B Reduced-Form Evidence

Table A.1 reports a multinomial logit of the three outcomes on realized suspense and surprise, the commercial flag, and viewer characteristics, estimated separately for viewers who were not tuned in and who were tuned in in the previous minute, with game and inning fixed effects. Among viewers already tuned in, both suspense and surprise raise the probability of active attention relative to not being tuned in, and commercials lower the probabilities of both passive and active attention. Among viewers who were not tuned in, realized suspense predicts entry into active attention; as discussed in Section 4, this association reflects viewers' expectations about the game, which the structural model captures through expected suspense.

Figure A.1: Suspense and Surprise by Inning



Note: Averages and 95% confidence intervals of at-bat level suspense (left) and surprise (right) by inning.

Appendix C Specification Ladder

Table A.2 reports the fit of the sequence of specifications that leads to the model of Section 4. Each row adds the stated element to the previous row. The largest improvements come from choice persistence (J3) and from the persistence of attention in the final specification (J10); the fit is evaluated on the same 377,529 viewer-minutes for all specifications.

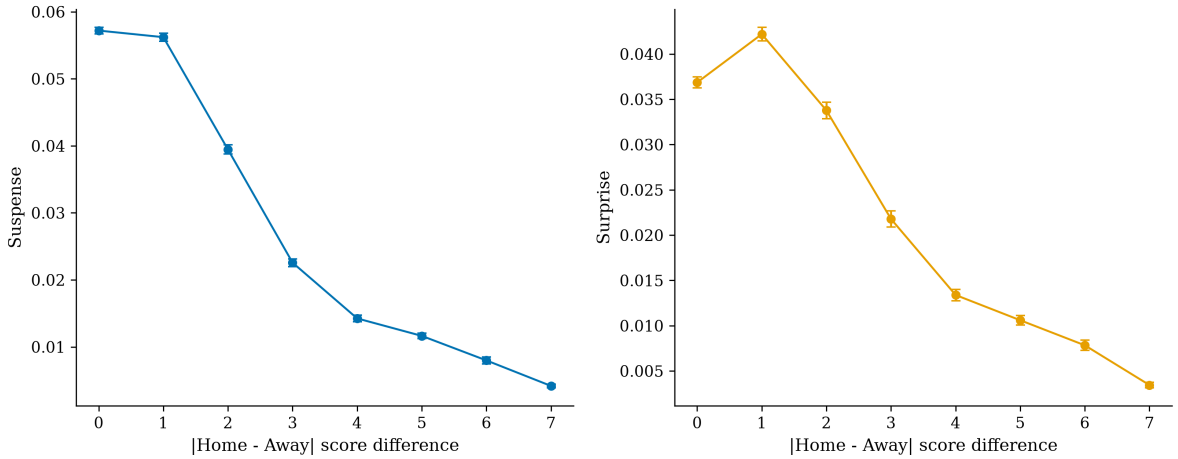
Appendix D Details of the Counterfactual Exercises

Candidate schedules. A block may start at any observed block start or within one minute of a half-inning change. Each game keeps its observed multiset of block durations; blocks may not overlap and must be separated by at least one minute of play.

One-step optimum. For each game we solve a mixed-integer linear program that assigns blocks to start minutes to maximize expected active attention, evaluated with the observed choice histories and beliefs held fixed, so that the schedule affects attention only through the current and lagged commercial indicators.

Recursive simulation and optimization. The recursive simulator draws tune-in and attention minute by minute from the estimated model. Simulated active attention to game content updates the viewer's last observed state, information recency, and subsequent beliefs; simulated choices become the next minute's lagged choices. Realized content is held at its pre-break value during each schedule's own breaks. The tune-in and attention intercepts are calibrated so that the observed schedule reproduces the observed numbers of tuned-in and active minutes. All schedules are evaluated with the same 500 sets of random draws. The recursive optimum is found by a local search that starts from the one-step optimum, screens one-block relocations and duration swaps with the one-step objective, and evaluates the most promising and a random set of neighbors with the recursive simulator.

Figure A.2: Suspense and Surprise by Score Difference



Note: Averages and 95% confidence intervals of at-bat level suspense (left) and surprise (right) by the score difference.

Frontier and ad-attention schedules. For each game we build a library of candidate schedules by local searches that maximize weighted sums of total active attention and active attention during commercials. A multiple-choice integer program then selects one schedule per game to maximize total active attention subject to a minimum aggregate level of active attention during commercials (the frontier), or to maximize active attention during commercials subject to a maximum loss of attention to the game (the ad-attention schedules). Both selections are exact within the library.

Simple rules. Each rule solves an exact placement problem for a simple score under the same constraints: minimizing the total realized suspense (or surprise) shown during commercial minutes, the same with each block restricted to start within 15 minutes of an observed block start, or maximizing the number of commercial minutes in innings 7 and later.

Inference. Game-bootstrap intervals resample the 43 games with replacement (9,999 replications) and recompute the total effect from the game-level simulated means, holding coefficients and schedules fixed. The re-estimation bootstrap draws 50 game-cluster samples, re-estimates all coefficients and fixed effects in each, and re-evaluates the fixed schedules, keeping the intercept calibration of the full-sample model.

Appendix E Shortening the Game: Construction and Decomposition

This appendix gives the details behind the seven-inning counterfactual of Section 5.3.

Seven-inning win-probability surface. We implement the shortened game under the convention that a seven-inning game is played like a nine-inning game resumed in the third inning: we retain each game's innings three through nine and relabel them one through seven. Under this rule the game ends after the seventh inning, so the home-win probability surface of Section 3 no longer applies and must be recomputed with the terminal boundary at the end of the seventh. We do so by value iteration on the at-bat transition kernel: the home-win probability of a state is the probability-weighted average of the home-win probabilities of its successor states, with the three absorbing end

Table A.1: Multinomial Logit by Tune-in Status in the Previous Minute

	Not tuned in at $t-1$		Tuned in at $t-1$	
	Passive	Active	Passive	Active
Suspense	0.475*	0.847*	0.040	0.685**
	(0.264)	(0.491)	(0.249)	(0.279)
Surprise	0.228	−0.015	0.546**	0.701***
	(0.228)	(0.435)	(0.225)	(0.249)
Commercial	−0.061**	−0.045	−0.184***	−0.219***
	(0.028)	(0.051)	(0.025)	(0.028)
Age	0.000	0.005***	0.010***	0.015***
	(0.001)	(0.001)	(0.000)	(0.001)
Female	−0.019	−0.008	−0.113***	0.029
	(0.019)	(0.035)	(0.018)	(0.020)
Home fan	−0.004	0.120**	−0.083***	−0.010
	(0.033)	(0.058)	(0.031)	(0.035)
Away fan	0.297***	0.290***	−0.126**	−0.262***
	(0.056)	(0.094)	(0.054)	(0.062)
Observations	476,283		139,866	
Log likelihood	−73,206		−118,187	

Note: Base outcome: not tuned in. Columns 1–2: viewers not tuned in at $t-1$; columns 3–4: viewers tuned in at $t-1$. Game and inning fixed effects; heteroskedasticity-robust standard errors in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels.

states (home win, home loss, tie) fixed at one, zero, and one half, iterated to convergence. Suspense and surprise are then recomputed from this surface exactly as in Section 3. Rebuilding the surface with the original nine-inning boundary reproduces the win-probability model of Section 3 with a correlation of 0.99, which validates the recomputation. We then re-simulate every retained viewer-minute under the fitted model, resetting each viewer's lagged choices and information recency at the new first inning, and recompute beliefs from the shortened-game transition kernel.

Partial identification of the end-of-game hazard. One ingredient of the seven-inning surface is not identified from nine-inning data: the within-inning hazard that a game ends on a seventh-inning walk-off, since essentially no observed game ends in the seventh. The deterministic end rule, that a non-tied game is over after the bottom of the seventh, is identified; only the magnitude of this mid-inning hazard is not. We therefore treat the counterfactual as partially identified over that hazard and recompute it across a range of hazard magnitudes, from zero (the game ends only at the deterministic boundary) to twice the transplanted end-of-game hazard. The results are nearly invariant: mean suspense stays at 0.034–0.035, the per-minute change in active attention during gameplay ranges from −2.45 to −2.52 percent, and the total-attention change from −20.11 to −20.16 percent. The one object the data cannot pin down does not affect the conclusion, because the deterministic end rule, not the walk-off hazard, governs how much shorter the game is.

Channel decomposition. The per-minute change in active attention is the net of three channels, which we isolate by turning them on in sequence (gameplay active attention per minute, pooled across viewers). The observed nine-inning game gives 2.81 percent. Dropping innings one and two but holding each viewer's history and the content measures fixed *raises* it slightly to 2.83 percent, because the later innings that remain are watched somewhat more closely (the composition effect).

Table A.2: Specification Ladder

Spec.	Added to the previous specification	Joint log likelihood	Brier score
J0	Belief moments	−163,875.1	0.22166
J1	+ demographics and commercial	−163,345.9	0.22121
J2	+ game FE	−162,209.9	0.22034
J3	+ choice persistence	−68,915.3	0.08408
J4	+ hour FE	−68,844.4	0.08405
J5	+ information-recency FE	−69,104.2	0.08440
J6	+ inning FE	−69,086.0	0.08439
J7	+ commercial interactions	−68,990.5	0.08418
J8	+ home/away fandom	−69,002.5	0.08420
J9	Realized content in attention equation	−69,002.8	0.08420
J10	+ realized-content fandom interactions	−67,879.6	0.08328

Note: Joint log likelihood of the three outcomes and multinomial Brier score. J10 is the specification reported in Table 5; relative to J9 it adds the fandom interactions and lagged active and passive attention to the attention equation.

Adding the reset of each viewer's viewing history at the new first inning lowers it to 2.75 percent: because active attention is highly persistent (Table 5), the attention that early-inning viewing would have carried into the middle of the game is lost when the game begins in what was the third inning. Finally, adding the recomputed seven-inning suspense and surprise moves it only to 2.74 percent. The history reset, not the content concentration the shortening proposal relies on, accounts for essentially all of the per-minute decline; the intensive-margin content channel is negligible. A variant that instead retains innings one through seven is similar in the total-attention change and shows a larger per-minute decline, since it removes the more heavily watched late innings.

Appendix F State Dependence versus Unobserved Heterogeneity

The attention and tune-in equations include lagged choices, and their coefficients are large: being tuned in at $t-1$ raises the log odds of being tuned in by about 5.6, and active attention at $t-1$ raises the log odds of active attention by about 1.4 (Table 5). A standard concern is whether this persistence reflects genuine state dependence, that watching makes a viewer more likely to keep watching, or unobserved heterogeneity, that some viewers are simply high-attention types whose stable trait induces serial correlation the lagged choice then absorbs. The distinction matters for the counterfactuals, which reallocate when viewers are watching.

We assess this by estimating the model as a latent-type mixture. Each viewer is one of K types, with type-specific intercepts and content and fandom slopes in both equations, while the persistence terms, demographics, and fixed effects are common across types; the type probabilities and all coefficients are estimated jointly by maximum likelihood. If the persistence were an artifact of unobserved heterogeneity, the type-specific intercepts would absorb the stable viewer traits and the common persistence coefficients would fall. Table A.3 reports the result for $K = 1$ (the baseline model of Section 4), $K = 2$, and $K = 3$.

Two or three types improve fit (the BIC falls from 137,261 to 135,563), so the data do contain viewer heterogeneity; the two-type model splits viewers into a minority low-attention segment (22 percent) and a majority segment (78 percent). The test therefore has power to detect unobserved heterogeneity if it were masquerading as state dependence. It is not: the persistence coefficients barely move as types are added, changing by at most about five percent, and the dominant tune-in

Table A.3: Latent-Type Mixture: Persistence and Fit

	One type (baseline)	Two types	Three types	Change $1 \to 3$
Persistence (common across types)				
Tuned in at $t-1$	5.56	5.42	5.43	−2.3%
Active at $t-1$	1.41	1.42	1.33	−5.5%
Passive at $t-1$	−1.49	−1.46	−1.41	+4.8%
Model fit				
Log likelihood	−68,092	−67,483	−67,171	
Parameters	164	175	186	
BIC (viewers)	137,261	136,114	135,563	
Type shares	1.00	0.22, 0.78	0.25, 0.66, 0.09	

Note: Joint latent-type mixture over the tune-in and attention equations; each viewer is one of K types. Persistence (lagged-choice) coefficients, demographics, and fixed effects are common across types; intercepts, content, and fandom slopes are type-specific. Estimated by maximum likelihood on the 380,092 tune-in and 48,298 attention viewer-minutes. "BIC (viewers)" uses the number of viewers (709) as the sample size. "Change $1 \to 3$" is the percent change in the common persistence coefficient from the one-type baseline to the three-type model.

persistence (5.56) changes by 2.3 percent. The content and fandom coefficients become type-specific and, split across 43 game clusters, individually less precise, but they stay sign-consistent: the neutral surprise response in attention remains positive, the home-fan suspense response remains positive, and the away-fan surprise response remains strongly negative in every type. We conclude that the estimated persistence reflects genuine state dependence rather than unobserved viewer heterogeneity, so the counterfactuals, which operate through that persistence, are not driven by a composition artifact.

Appendix G A Dynamic Model

The model in the main text is static: a viewer responds to the suspense and surprise in front of her but does not weigh how watching now changes what she will know, and value, later. Because suspense is by construction about the anticipation of future belief movements, it is worth asking whether a forward-looking viewer, who tunes in partly to be present when the game turns, fits the data better. This appendix estimates such a model and finds that it does not improve on the static one, which is why we keep the static specification in the body.

We extend the three-choice model to a finite-horizon dynamic discrete choice problem over the minutes of a game. Each minute the viewer chooses among not tuned in, passive, and active; the flow utilities are those of Section 4, augmented with two switching costs, a cost κ_T of entering the tuned-in state and a cost κ_A of entering active attention, which capture the persistence that the lagged choices carry in the static model. The viewer's information evolves exactly as in the main model (the state is revealed only by active watching), and she forms expectations over future suspense, surprise, and commercial breaks using the same transition kernel and empirical commercial hazard. She discounts future utility at a fixed rate $\delta = 0.999$ per minute. We solve the problem by backward induction and estimate it by maximum likelihood; at $\delta = 0$ the model reduces exactly to the static multinomial logit, which we use as a nesting check.

Table A.4 reports the estimates. The switching costs are large and precisely estimated: $\kappa_T = 5.32$ and $\kappa_A = 2.90$, which is the structural counterpart of the strong persistence in the static model,

Table A.4: Dynamic Model Estimates

	Estimate	Game-clustered SE
Switching costs		
Tune-in cost κ_T	5.32^{***}	0.05
Attention cost κ_A	2.90^{***}	0.06
Content, flow utility (passive)		
Suspense	-13.97^{***}	0.73
Surprise	11.17^{***}	0.56
Commercial	-0.04	0.03
Content, flow utility (active)		
Suspense	-12.28^{***}	0.73
Surprise	9.55^{***}	0.65
Commercial	-0.06	0.05
Discount factor δ (fixed)	0.999	
Viewer-minutes	620,853	

Note: Maximum-likelihood estimates of the finite-horizon dynamic three-choice model, in flow-utility units (not directly comparable to the logit coefficients of Table 5, since the dynamic choice values also include continuation values). Switching costs κ_T and κ_A are the costs of entering the tuned-in and the active state. Standard errors clustered by game. δ is fixed. *** denotes significance at the 1% level.

and the content coefficients keep the signs they have in the static specification. The dynamic model fits far better in sample than its static special case, as such a comparison almost always does, so the relevant test is out of sample.

We compare the dynamic model with the matched static model ($\delta = 0$, same flow-utility specification) by their ability to predict each game, scoring every viewer-minute by its log loss under each model at that model's own estimates. Table A.5 reports the result. The static model predicts better: its pooled log loss is lower, and it predicts better in 45 of the 46 games. Making the viewer forward-looking, in short, does not help explain the data, so we report the static model, which is simpler and whose coefficients map directly to choice probabilities.[14]

Table A.5: Static versus Dynamic: Out-of-Sample Prediction

	Static ($\delta = 0$)	Dynamic ($\delta = 0.999$)
Pooled mean log loss	0.275	0.305
Games predicted better (of 46)	45	1
Mean per-game log loss, dynamic − static	$+0.029$	$[-0.007, +0.078]$

Note: Mean per-minute log loss (lower is better) under the matched static and dynamic models, each scored at its own full-sample estimates. "Games predicted better" counts the 46 games by which model attains the lower log loss. The bracketed range is the minimum and maximum per-game difference across games.

[14] Because the two models are scored at their full-sample estimates rather than refit on training games, this is a fixed-parameter predictive comparison rather than leave-one-out cross-validation; the comparison is informative because the in-sample advantage of the richer model is symmetric across the two. The comparison shows that a forward-looking model does not predict better here, not that viewers are myopic in a deeper sense.

Appendix H Information Assumptions and History Controls

The information assumption of our main specification, I0, lets viewers learn the game state through active attention: expected content enters tune-in and realized content enters attention conditional on tuning in. We assess its sensitivity against two alternatives. Under I1, passive viewing of gameplay can also reveal the state. Under I2, a viewer who actively watched gameplay at $t-1$ uses realized content at t in both the tune-in and attention equations; all other viewers use expected content from their latest informative state under the no-cutoff belief rule. For viewers who have never observed the game, the empirical prior supplies the initial belief. I2 grants advance information about the current play by assumption, which is why we treat it as a robustness check rather than the main specification. The same hybrid content regressor enters both equations, so I2 has its own tune-in coefficients rather than sharing I0's equation. Results produced under the earlier expected-content-only version are labeled legacy I2 below.

We compare these assumptions using identical samples and controls. The richest specification, R4, includes both active-information recency and tuned-viewing recency, fine categories for the viewing or absence spell ending at $t-1$, and indicators for censored histories. Histories use the full observed timeline before selecting estimation observations. There are 380,092 tune-in observations across 46 games and 48,298 attention observations across 43 games.

For the I2 specification, let $D_{i,t-1}=1$ when viewer i actively watched gameplay in the preceding minute. The hybrid content measures are

$$\widetilde{S}^{I2}_{it} = D_{i,t-1}S_t + (1-D_{i,t-1})\bar{S}_{it}, \qquad \widetilde{U}^{I2}_{it} = D_{i,t-1}|U_t| + (1-D_{i,t-1})\bar{U}_{it}. \tag{6}$$

The current I2 tune-in and attention indices are

$$\begin{aligned}
V^{T,I2}_{it} &= \beta^T_S \widetilde{S}^{I2}_{it} + \beta^T_U \widetilde{U}^{I2}_{it} && (7)\\
&\quad + \gamma^T_0 CM_t + \gamma^T_1 CM_{t-1} + x_i'\theta^T + \phi^T_H H_{ig} + \phi^T_R R_{ig} + \delta^T_{it}, && (8)\\
V^{A,I2}_{it} &= (\beta_S + \beta^H_S H_{ig} + \beta^R_S R_{ig})\widetilde{S}^{I2}_{it} && (9)\\
&\quad + (\beta_U + \beta^H_U H_{ig} + \beta^R_U R_{ig})\widetilde{U}^{I2}_{it} && (10)\\
&\quad + \lambda_A A_{i,t-1} + \lambda_P P_{i,t-1} + \gamma^A_0 CM_t + \gamma^A_1 CM_{t-1} && (11)\\
&\quad + x_i'\theta^A + \phi^A_H H_{ig} + \phi^A_R R_{ig} + \delta^A_{it}. && (12)
\end{aligned}$$

Here δ^T_{it} and δ^A_{it} include game, clock-hour, inning, both recency dimensions, fine viewing/absence spells, and censoring controls. The separate lagged tune-in indicator is omitted: tuned-viewing recency fixed effects absorb it in R4. The estimates below therefore report no standalone lagged tune-in coefficient.

Under current I2, tune-in suspense is positive (1.383, $p=0.016$), while tune-in surprise is near zero (-0.004, $p=0.994$). For neutral viewers, both suspense and surprise raise active attention (1.588, $p=0.011$; and 1.075, $p=0.013$). The content and fandom interaction terms are jointly significant, though the group-specific interaction estimates are less precise. The hybrid regressor uses current realized content on 2.36 percent of tune-in rows and 16.12 percent of attention rows; most identifying content variation therefore comes from the expected-content component. The attention likelihood improves modestly relative to I0, but this is an in-sample comparison and does not identify the assumed timing of information.

Under no-cutoff R4, suspense is positive in the I0 tune-in equation ($p=0.019$) and in I2 ($p=0.016$); the I2 surprise coefficient is near zero ($p=0.994$). For neutral viewers, surprise is positive in attention under I0 and I1 ($p<0.001$) and under I2 ($p=0.013$). Because I2 grants

Table A.6: Current I2 tune-in coefficients, no-cutoff R4

Term	Estimate	Game-clustered SE
Hybrid suspense	1.383**	0.576
Hybrid surprise	-0.004	0.587
Commercial at t	-0.480***	0.041
Commercial at $t-1$	0.052	0.045
Age	0.002***	0.000
Female	-0.051**	0.020
Home fan	0.039	0.035
Away fan	0.038	0.043
Observations	380,092	
Games	46	
Log likelihood	-47,898.94	

Note: $p < 0.10$, $p < 0.05$, and $p < 0.01$ are denoted by *, **, and ***, respectively. The hybrid content measure uses current realized content after active gameplay at $t-1$ and expected content otherwise. Tune-in content terms are jointly significant ($p = 0.0505$).

Table A.7: Current I2 attention coefficients, no-cutoff R4

Term	Estimate	Game-clustered SE
Hybrid suspense, neutral	1.588**	0.624
Hybrid surprise, neutral	1.075**	0.432
Suspense × home fan	3.840***	1.460
Surprise × home fan	-1.842	1.756
Suspense × away fan	-2.218	3.670
Surprise × away fan	-14.716***	5.339
Commercial at t	-0.340***	0.058
Commercial at $t-1$	-0.104*	0.054
Active at $t-1$	1.403***	0.162
Passive at $t-1$	-1.424***	0.101
Age	0.003***	0.001
Female	0.051	0.043
Home fan	-0.103	0.109
Away fan	0.441**	0.221
Observations	48,298	
Games	43	
Log likelihood	-16,550.24	

Note: Content terms are jointly significant ($p = 0.0000144$); the four fandom interactions are jointly significant ($p = 0.00211$). The content coefficients for neutral viewers are the reference slopes.

Table A.8: Content coefficients under the preferred no-cutoff specification (R4)

Tune-in equation	I0	I1	I2
Content suspense	2.042	1.556	1.383
	(0.871)	(0.967)	(0.576)
Content surprise	0.320	1.250	-0.004
	(1.330)	(1.151)	(0.587)
Attention equation	**I0**	**I1**	**I2**
Suspense	0.610	0.717	1.588
	(0.443)	(0.463)	(0.624)
Surprise	0.867	0.896	1.075
	(0.230)	(0.258)	(0.432)

Note: Game-clustered standard errors in parentheses. Attention coefficients describe neutral viewers; all models include fandom interactions. I0 and I1 use realized gameplay content in attention. I2 uses current realized content after active gameplay at $t-1$ and expected content otherwise, in both equations.

advance information by assumption, these coefficients do not describe what a viewer can know before the play occurs. Legacy expected-content I2 attention estimates are imprecise: their 95% intervals are approximately $[-0.257, 3.655]$ for suspense and $[-0.819, 4.905]$ for surprise. An encompassing attention equation with both realized and expected content blocks was estimated under the original 15-minute reset; its conventional game-clustered Wald tests gave $p = 0.0016$ and $p = 0.0105$.

Belief-memory cutoff sensitivity. Our preferred belief rule carries the last informative state forward for every observed gap. Viewers who have not yet watched still receive the empirical prior at the current point of play. The cutoff comparison in Table A.9 reports the legacy expected-content-only I2. It compares 15- and 30-minute resets with no cutoff while holding R4 controls, samples, histories, and recency bins fixed. Under legacy I2, tune-in rows match I0 and are represented by the I0 rows in the table. I0's attention estimates do not change with the belief-memory rule. I1 uses realized gameplay content in attention and expected content during commercials. Legacy I2 uses expected content throughout.

Table A.9: Legacy expected-content I2: coefficients by belief-memory cutoff, common R4 controls

Equation	Model	15-minute reset	30-minute reset	No cutoff
Tune-in	I0 suspense	1.917 (0.929)	0.651 (1.036)	2.042 (0.871)
	I0 surprise	-0.459 (1.257)	1.585 (1.271)	0.320 (1.330)
	I1 suspense	0.824 (0.921)	-0.365 (0.831)	1.556 (0.967)
	I1 surprise	1.600 (0.984)	3.992 (0.872)	1.250 (1.151)
Attention	I0 suspense	0.610 (0.443)	0.610 (0.443)	0.610 (0.443)
	I0 surprise	0.867 (0.230)	0.867 (0.230)	0.867 (0.230)
	I1 suspense	0.739 (0.466)	0.733 (0.469)	0.717 (0.463)
	I1 surprise	0.874 (0.255)	0.884 (0.255)	0.896 (0.258)
	Legacy I2 suspense	1.373 (1.037)	0.673 (0.955)	1.699 (0.998)
	Legacy I2 surprise	1.617 (1.475)	3.113 (1.213)	2.043 (1.460)

Note: Game-clustered standard errors in parentheses. Attention coefficients are for neutral viewers; all attention models include fandom interactions. Every specification uses the same samples and R4 recency, spell, and censoring controls.

For legacy expected-content I2, attention likelihood is highest under no cutoff ($-16{,}556.198$, compared with $-16{,}560.513$ under 15 minutes); I1's likelihood is marginally higher as well ($-16{,}554.301$

versus −16,554.492). I0 is identical across cutoffs. These descriptive comparisons do not select an information assumption. Legacy-I2 coefficients move across cutoffs, so estimates remain sensitive to state-anchor duration. No cutoff propagates beliefs for gaps up to 274 minutes in this sample. The extension retains absorbing game-end states without conditioning on the viewer knowing that the game remains in progress, and the empirical event clock has fewer supporting games at long gaps. We retain no cutoff as the preferred belief rule and the 15- and 30-minute results as sensitivity checks. The current I2 estimates and policy counterfactuals use the no-cutoff rule.

Few-cluster fandom inference. Conventional group slopes should be read alongside inference based on the 43 game clusters. A null-imposed efficient-score multiplier test of the four fandom interactions gives baseline $p = 0.0447$ with Rademacher weights and $p = 0.0466$ with Webb weights; the corresponding R4 values are 0.0310 and 0.0325. This test uses restricted-model scores and a fixed score covariance, without refitting bootstrap draws. A complementary unrestricted likelihood bootstrap, refitting all coefficients with positive exponential game weights in 999 draws, gives centered Wald $p = 0.009$ at baseline and $p = 0.005$ under R4. All draws converge. This second calculation uses the original clustered covariance as its Wald metric and is not a null-imposed wild bootstrap. Both exercises support joint heterogeneity, while individual fan slopes remain sensitive to the bootstrap method.

Legacy-I2 policy sensitivity. Earlier policy comparisons used expected content throughout and a 15-minute belief reset. In a 200-draw comparison, the total-attention schedule raised active attention by 2.16% and ad attention by 7.15% under that legacy specification. These results remain a benchmark and do not describe the current I2 assumption.

Table A.10 extends the earlier comparison to R4, updating history controls recursively under each information assumption. The legacy-I2 column retains its original expected-content rule and 15-minute belief reset.

Table A.10: Legacy expected-content I2 policy sensitivity with common recency and spell controls (R4)

Fixed I0-selected schedule	I0	I1	Legacy I2
Total active attention			
Recursive schedule	+2.90%	+2.99%	+2.85%
Pareto schedule	+2.90%	+2.73%	+2.77%
Ad-attention schedule	+1.22%	+1.09%	+1.20%
Active attention during ads			
Recursive schedule	-20.33%	-20.04%	-19.57%
Pareto schedule	+1.53%	+1.18%	+0.48%
Ad-attention schedule	+8.92%	+8.80%	+7.93%

Note: Percent changes relative to each model's simulated observed schedule. Both recencies, fine viewing/absence spells, and censoring controls update along simulated paths. We use 200 paired common-random draws and 43 games with model-specific observed-schedule calibration. Schedules were originally selected under I0 and are held fixed across the columns; the legacy-I2 column uses expected content and the original 15-minute reset. The labels do not denote R4 optima. Game-bootstrap intervals condition on fitted parameters and schedules.

The recursive schedule raises total active attention by 2.85–2.99% across the three R4 models. The ad-attention schedule raises ad attention by 7.93–8.92%. These patterns survive the information comparison. The Pareto schedule's ad-attention intervals include zero under all three models, so positive point estimates do not guarantee its original no-loss constraint under R4.

We also let each R4 model select a total-attention schedule separately for each game from the existing observed, recursive, Pareto, and ad-attention schedules. The legacy expected-content I2 results in this finite-menu exercise use the original 15-minute reset and separate simulation seeds; they do not select a new I2 schedule or account for parameter uncertainty.

Recursive policy evaluation under I2. We now evaluate the existing policy schedules using the current I2 estimates, recursively updated viewer histories, and the no-cutoff belief rule. Each schedule was selected under an earlier model and is held fixed; this is not an I2-specific re-optimization. The model is calibrated to the observed schedule, then the resulting common intercept offsets are held fixed across all policies. Table A.11 reports percent changes relative to I2's simulated observed schedule. Intervals resample the 43 games and condition on the fitted coefficients and schedules.

Table A.11: Recursive policy evaluation under current I2

Existing policy schedule	Total attention	Attention to ads	Attention to game
Recursive optimum	+3.20% [1.68, 4.99]	−20.06% [-31.57, -10.46]	+6.15% [3.27, 9.55]
Pareto maximum	+3.99% [2.18, 6.11]	−10.94% [-20.81, -2.83]	+5.88% [3.01, 9.32]
Ad attention held at baseline	+2.89% [1.50, 4.60]	+0.40% [-8.39, 9.04]	+3.20% [1.27, 5.82]
Ad-attention optimum	+1.42% [0.47, 2.67]	+8.43% [2.63, 15.89]	+0.53% [-0.54, 1.81]
Low-suspense rule	+3.25% [1.71, 5.13]	−7.84% [-15.18, -1.52]	+4.65% [2.40, 7.40]
Protect game attention	+1.85% [0.94, 3.09]	+7.97% [2.19, 15.46]	+1.07% [0.06, 2.29]

Note: Entries are percent changes from the simulated observed schedule; brackets show 95% game-bootstrap intervals based on 9,999 resamples of the 43 games. Simulation uses 200 paired draws per game and policy. Intervals condition on point estimates and schedules; Monte Carlo standard errors are reported in the replication output. Existing schedules were not re-optimized under I2.

The recursive and Pareto maximum schedules raise total and game attention but lower attention to ads. The ad-attention optimum raises ad attention by 8.43%, while its estimated change in game attention is small and its interval spans zero. The schedule designed to protect game attention raises both ad and game attention in this evaluation. These are conditional predictions under I2's advance-information assumption, not causal estimates of what viewers know before a play.

Timing and latent types. A second-level event pilot suggests a delayed attention response around 20 seconds after recorded pitch timestamps. These timestamps have not been independently aligned to broadcast outcome revelation, and the pilot examines several lags, so it cannot identify prior information or reject I2. In separate held-out viewer-game validation, adding type-specific content slopes to a two-type model with type-specific histories produces no clear prediction gain. With training-only posteriors, the I0 held-out log-score gain is 2.32 (conditional game-bootstrap interval $[-2.93, 7.83]$), and the I2 gain is -1.33 ($[-4.42, 1.13]$). These intervals condition on fitted training parameters. We retain latent content types as an exploratory extension.

Holding the viewer cohort fixed across all four event windows preserves the exploratory delayed pattern: at 20 seconds, the shock coefficient is 0.234 (game-clustered SE 0.071, $p = 0.0018$) across 1,384 same-minute events in 45 games. The pre-event placebo is insignificant ($p = 0.317$). This cohort consists of viewers observed watching continuously in the windows and therefore selects on persistent viewing; it does not identify a population causal effect. Independent broadcast-clock calibration remains necessary.

Two additional viewer-game splits show that the latent-type conclusion depends on the split. With seed 104, typed content improves I0 held-out log score by 77.31, with a conditional interval

from 26.01 to 141.38. It worsens I2 by 137.53, with a gain interval from -252.57 to -38.79. Under seed 73, the I0 raw log-score gain is sensitive to extreme predicted probabilities: it falls from 179.80 with a 10^{-12} probability floor to 9.40 with a 10^{-4} floor. Its Brier-score interval spans zero. The seed-104 Brier scores corroborate improvement for I0 and deterioration for I2. We do not pool these overlapping splits as independent experiments. The evidence does not yet establish stable content-response types, and a full bootstrap of mixture training uncertainty remains outstanding.