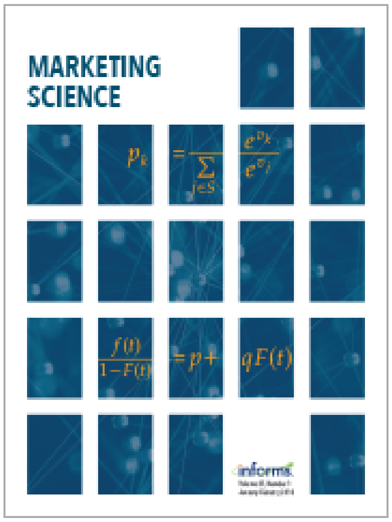


Marketing Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

How Viewer Tuning, Presence, and Attention Respond to Ad Content and Predict Brand Search Lift

Matthew McGranaghan, Jura Liaukonyte, Kenneth C. Wilbur

informs
http://pubsonline.informs.org/journal/mksc

MARKETING SCIENCE
Articles in Advance, pp. 1–23
ISSN 0732-2399 (print), ISSN 1526-548X (online)

How Viewer Tuning, Presence, and Attention Respond to Ad Content and Predict Brand Search Lift

Matthew McGranaghan,[a] Jura Liaukonyte,[b] Kenneth C. Wilbur[c]

[a] Alfred Lerner College of Business and Economics, University of Delaware, Newark, Delaware 19716; [b] SC Johnson College of Business, Cornell University, Ithaca, New York 14853; [c] Rady School of Management, University of California, San Diego, La Jolla, California 92093

Contact: mmcgran@udel.edu, https://orcid.org/0000-0002-6747-9867 (MM); jurate@cornell.edu, https://orcid.org/0000-0002-9820-8832 (JL); kcwilbur@ucsd.edu, https://orcid.org/0000-0002-1227-8750 (KCW)

Received: March 29, 2021
Revised: September 3, 2021; November 2, 2021
Accepted: November 4, 2021
Published Online in Articles in Advance: February 9, 2022

https://doi.org/10.1287/mksc.2021.1344

Abstract. New technology measures TV viewer tuning, presence, and attention, enabling the first distinctions between TV ad viewability and actual ad viewing. We compare new and traditional viewing metrics to evaluate the new metrics' utility to advertisers. We find that 30% of TV ads play to empty rooms. We then use broadcast networks' verifiably quasi-random ordering of ads within commercial breaks to estimate causal effects of ads on new viewing metrics among four million advertising exposures. We measure ad metadata and machine-code content features for 6,650 frequent ad videos. We find that recreational product ads preserve audience tuning and presence. Prescription drug advertisements decrease tuning and presence, more so for drugs that treat more prevalent and severe conditions. We also investigate whether new viewing data can inform advertiser objectives, finding that attention helps predict brand search lift after ads.

History: Olivier Toubia served as the senior editor and Hema Yoganarasimhan served as associate editor for this article.

Conflict of Interest Statement: The authors have no funding or conflicts of interest to report.

Supplemental Material: Data and the online appendix are available at https://doi.org/10.1287/mksc.2021.1344.

Keywords: advertising • audience measurement • attention • ad content • ad avoidance • ad response

1. Introduction

Advertising-supported TV, which in the United States commanded $70 billion in advertising revenue in 2019 (eMarketer 2019), operates on the basis that consumers "pay" attention to ads. However, this exchange of attention to advertising for attention to content is not subject to an explicit agreement; that is, consumers can stop paying attention to ads, leave the room, or change the channel—behaviors that influence the value of advertising.

An established industry standard to measure viewership relies on sampling tuned-in TVs to obtain estimated audience sizes. Recently, the ability of this traditional audience measurement to accurately track viewers' behavior has been called into question by an industry watchdog, which temporarily suspended accreditation of a national TV ratings service (Steinberg 2021). Fortunately, new technology allows for more detailed and granular measurement, including tracking the moments viewers are present in the room and paying attention to the screen.

This paper asks three questions about TV advertising viewing: How do traditional tuning data compare with new TV viewing metrics? How do new viewing metrics respond to ad content? Do new viewing metrics predict ad response better than traditional ones? The answers can inform advertisers about the potential value of moving to a new standard of audience measurement.

We investigate viewing behavior measured using cameras, microphones, and algorithms in a paid sample of 1,155 consenting households (3,659 viewers). *Tuning* is measured by comparing television audio to a database of known programs. Individual viewer *presence* is measured using person-detection and facial-recognition algorithms. *Attention* is measured as the co-occurrence of eyes-open and eyes-on-screen inferences.

All three viewing metrics are measured in situ, passively and continuously.

These measures provide a more granular picture of how viewers engage with TV ads. For instance, viewer presence detection distinguishes true ad exposures from ads that air to empty rooms. In our sample, viewers are absent from the room during 30% of the ads that play on their TV during active viewing sessions, and viewers are about four times more likely to leave the room during an ad than to tune away. We also see stark differences in ad viewing behaviors across common ad targeting criteria—channel, daypart, program genre, gender, and age. For example, during primetime, viewers are more likely to tune away and more likely to pay more attention to ads, relative to other dayparts.

We investigate whether there is a relationship between these new measures and advertising content by combining viewer behavior data with ad metadata and machine-coded ad content features for 6,650 frequent ad videos. We verify and exploit broadcast networks' practice of quasi-random ordering of ads within breaks to identify causal effects of ad content on tuning, presence, and attention. Our results suggest that using a naive model inflates the effects of advertising content, and that most ad content effects are small. However, there are a few notable patterns in the estimated ad content effects: viewer tuning and presence fall less during recreational product ads; prescription drug ads reduce tuning and presence more than average, more so for drugs that treat severe and prevalent conditions; and attention decreases across the first three ad slots in a break and falls with ad duration.

Finally, we investigate whether the new viewing metrics predict brand search lift after TV ads better than traditional metrics. We find that attention helps predict online search response to ads, as does distinguishing ad exposures from viewable ads. These new metrics therefore provide new and relevant information that can help inform advertiser objectives.

The new measures and subsequent ad content results have implications for advertisers, ad-selling platforms, and viewers. For advertisers, these granular outcome data can inform campaign decisions and media buying, as well as serve as an intermediate ad performance metric. For ad-selling platforms, these findings suggest that ad content effects on viewing behaviors can be used to select, target, order, and price ads to optimize ad delivery. For example, a platform can restrict or subsidize ad content to help preserve viewer presence and attention. For viewers, presence- and attention-preserving ad market designs can have implications for satisfaction with user experience and long-term usage of a platform. Incorporating these metrics in ad market design could be a rare win-win-win for advertisers, platforms, and consumers.

Next, we discuss the relevant advertising literature. Section 2 describes the new viewing and ad content metrics. Section 3 specifies the model and causal identification strategy. Section 4 presents the ad content findings and then explains the drug results using treated condition attributes. Section 5 relates viewing metrics to online search. Section 6 concludes with limitations and possible extensions.

1.1. Relationship to Previous Literature

Advertising studies usually balance viewing behavior measurement quality and sample size. For example, many papers study how ads change TV tuning in large field samples (Danaher 1995, Shachar and Emerson 2000, Goettler and Shachar 2001, Wilbur 2008, Schweidel and Kent 2010, Swaminathan and Kent 2013, Wilbur 2016). A distinct literature studies how ads change viewer attention and emotion in small laboratory samples (Zhang et al. 2009, Teixeira et al. 2012, Liu et al. 2018). A third literature studies ad viewing in small-scale ethnographic samples (Jayasinghe and Ritson 2013, Voorveld and Viswanathan 2015).

The current paper is likely the first to combine laboratory-like ad viewing metrics with large field samples. We know of one other paper that studies similar ad viewing data. Liu et al. (2021) quantify suspense and surprise during baseball games and find that in-game suspense decreases consumer attention during commercials, whereas in-game surprise enhances ad attention.

We join a growing number of studies that combine data on TV ad avoidance and ad response. The first we know of was Zufryden et al. (1993), who found that households' TV ad "zapping" decisions correlate with their future packaged good purchases. Siddarth and Chattopadhyay (1998) and Tuchman et al. (2018) used household purchase data to predict TV ad avoidance, finding that consumers are less likely to avoid ads for brands that they have previously bought. Bronnenberg et al. (2010) analyzed a field experiment that treated households with free digital video recorders, finding a tight null treatment effect on packaged good purchases. Deng and Mela (2018) combined device-level ad avoidance and sales data to study the consequences of microtargeted TV advertising. We contribute to this literature by quantifying how new viewing metrics respond to ad content and predict brand search lift.

2. Data and Descriptive Results

New metrics require careful definition, description, and comparison with traditional metrics. We describe the viewing data, introduce the ad data, and then describe their covariation.

2.1. Ad Viewing

Viewer tuning, presence, and attention data are provided by TVision Insights, an analytics firm founded to modernize television audience measurements. The data cover 3,659 viewers in a panel of 1,155 consenting households between July 2016 and June 2017.

2.1.1. Measurement. TVision installs cameras and microphones on each household's primary TV. Initial setup includes training facial-recognition algorithms on each household member. Infrared sensors measure depth and aid detection in low light conditions. Image data are processed in real time at the frame level five to six times per second on average. Images are not stored or transmitted outside of the home. TVision combines audio data with industry-standard automated content recognition services to measure television tuning, that is, the television network and timestamp of the audio stream.

TVision uses person-detection algorithms to identify human bodies—sets of heads, shoulders, and arms—in the cameras' field of view. Person-detection technology is similar to real-time face and body recognition algorithms used in smartphone apps, for example, Instagram filters. For each face, the software either identifies the household member or assigns a unique guest identifier. *Presence* is the detection and recognition of a particular viewer in the room.

TVision software measures when viewers' eyes are open and infers head orientation based on the relative sizes, positions, and angles of facial features. *Attention* is the co-occurrence of eyes-open and eyes-on-screen inferences.[1]

At the time the sample data were produced, TVision equipment measured tuning, presence, and attention continuously and then sampled one measurement for each viewer in each ad second. The data provided to us report average behaviors across viewer seconds within each viewer and ad insertion. So, viewing behaviors within 30-second ad exposures are based on 30 underlying measurements per viewer.

2.1.2. Comparisons to Extant Advertising Audience Measurements. Nielsen and TVision both measure tuning continuously and passively. However, two major advantages of TVision data are passive, continuous presence measurement and attention data.

Traditional television audience measurements are based on digital devices—mostly smart TVs and set-top boxes—and Nielsen People Meters. Digital devices measure tuning passively and continuously in millions of households, but do not measure which household members are watching at which times, or whether anyone is watching at all. People Meters measure tuning passively and continuously in representative samples of tens of thousands of households, and they additionally measure viewer presence in an intrusive and intermittent fashion. People Meters use a red light to prompt Nielsen panelists to log in on a special remote control at the start of each viewing session and once every 15 to 45 minutes thereafter. Nielsen combines viewer presence data with tuning data to determine audience demographics and infer when viewing sessions may have concluded.

Media buyers have long known that Nielsen audience estimates overstate advertisement audiences. Ephron (2006) argued that [C]ommercial-minute data . . . show losses of audience of about 2 to 10% during commercials compared with programs . . . Researchers, who read the fine print, qualify a Nielsen commercial exposure as '*an opportunity to see*' a commercial. And given the opportunity, it's obvious the probability is a lot less than one. So the Nielsen commercial-minute audience is an overstatement of people seeing commercials." In contrast, TVision's passive and continuous presence measurements avoid disrupting natural viewing behaviors and distinguish opportunities to see from actual ad exposures.

In what follows, we define an *opportunity to see (OTS)* as a viewer's television tuned to an ad insertion for at least two seconds, for any commercial break in which the viewer is present for at least two seconds in the first ad slot of the break. Selecting viewers present at the start of the break removes inactive viewing sessions from the sample. The two-second threshold is inspired by the Media Rating Council's definition of a viewable impression, in which 100% of an ad's pixels play on a screen for at least two seconds (Knauer 2019).

The definition of a viewable impression does not require a human to be exposed to the ad. Industry reports estimate that 10% to 30% of digital ad spend is lost to ad fraud, often because ads are served to machines instead of to humans (Gordon et al. 2021). For example, an analysis by the IAB Tech Laboratory indicated that only 59.8% of ad clicks could be confirmed as human traffic (Swant 2019). TVision presence data may offer the first passive, continuous human detection data in the history of video advertising.

We define an *ad exposure* as any OTS in which a viewer is detected as present for at least two seconds. An example can illustrate how ad OTS differ from ad exposures: Suppose a viewer watches a program that goes to an ad break. The break starts with a Coca-Cola ad, then a Geico ad, then follows with five other ads. The viewer leaves the room halfway through the Geico ad in the second slot and does not return until after the break ends. The viewer has had seven opportunities to see ads and two ad exposures (Coca-Cola and Geico).

The other major advantage of TVision data are attention measurement. TVision provides the first continuous, passive measurements of television viewers' ad attention in natural viewing environments. Viewer attention is becoming increasingly scarce as consumers increasingly use smartphones or tablets alongside television; attention measurements may help improve advertiser choices.

2.1.3. Ad Viewing Descriptives. Is tuning a reliable proxy for presence or attention? We compare the three behaviors in samples of ad OTS and ad exposures. The following graphics (Figures 1, 2, 3, and 4) focus on viewers with at least 50 ad exposures and commercial breaks on top-four broadcast networks between 7:00 a.m. and 1:00 a.m.[2]

Figure 1 compares densities of viewers' average tuning, presence, and attention behaviors during OTS and ad exposures. The average viewer's television remains tuned to 96.3% of viewable TV ad seconds. However, the average viewer remains present for only 54.6% of all ad seconds during OTS, with substantial heterogeneity in average viewer presence resulting in a 10–90th percentile range of 28.2%–76.7%. Further, the average viewer only pays attention to 7.7% of ad seconds during OTS. In fact, 7.2% of viewers disregard 99% or more of all viewable ad seconds on average. Ignoring the distinction between OTS and exposures overestimates ad viewing because 29.8% of the observations occur when ads play to empty rooms.

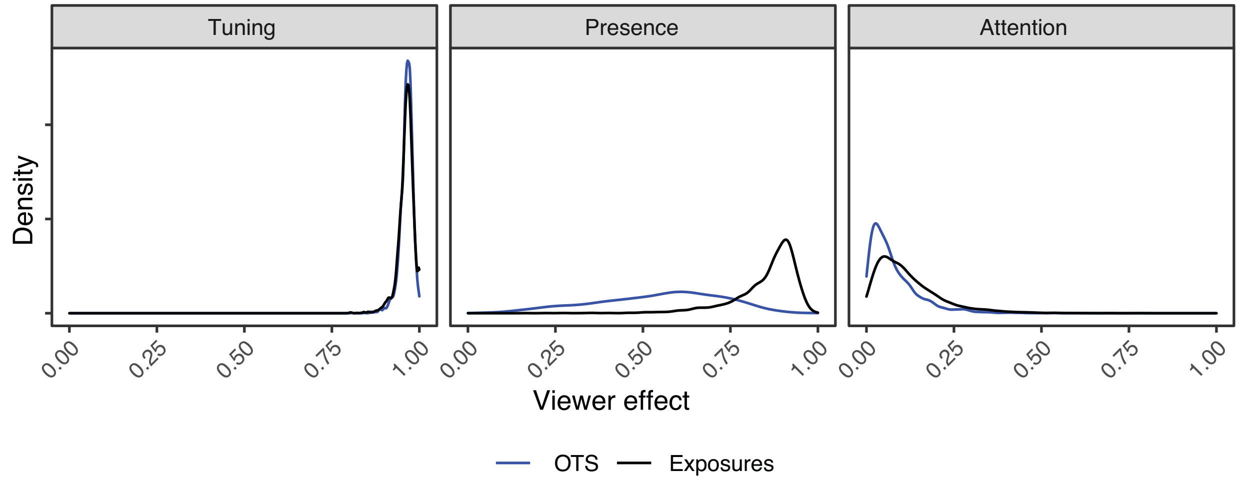
Comparing OTS densities to exposure densities, the tuning distribution changes little, with an average of 96.2% during ad exposures. However, average viewer presence increases from 54.6% to 85.3% and variation across viewers in average presence falls by about half after filtering out nonexposures. Average viewer attention increases from 7.7% to 11.7%. Only 3.1% of viewers disregard more than 99% of all ad seconds during exposures.

Figure 2 depicts covariation among viewers' average behaviors. Each point plots an individual viewer's average of two viewing behaviors among all ad exposures observed. All three panels show diffusion around strong central tendencies, indicating that the three behaviors are correlated yet still quite distinct. For example, within the subset of viewers who average 95% tuning, their average presence ranges from 57% to 99%. Within the subset of viewers who average 85% presence, their average attention ranges from 1% to about 41%. In sum, people engage in different ad viewing behaviors at quite different rates. Thus tuning is an incomplete proxy for presence or attention.

Figure 3 illustrates average tuning, presence, and attention within six age/gender groups in the OTS and exposure samples. Tuning varies minimally across demographics and samples. However, the OTS sample shows large differences in viewer presence across demographic groups, with older females showing the highest average presence at 67.3% and younger males showing the lowest average presence at 50.6%. However, the ad exposure sample shows muted variation with mean presence ranging from 85.8%–90.8% across groups. Therefore, viewers physically depart during ads (i.e., nonpresence) about three to four times more often than they change channels during ads (i.e., nontuning).

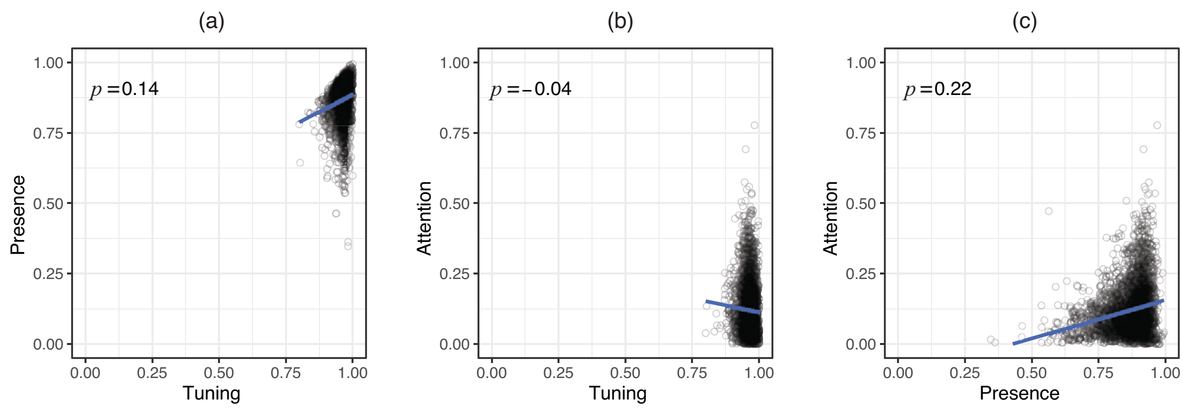
Like presence, ad attention increases with viewer age for both genders. However, unlike presence, removing nonexposures from the OTS sample does not change variation across groups much; instead, it mostly induces a level shift in mean attention. The level shifts imply that people leave the room during ads they are unlikely to have watched otherwise. Overall, patterns of tuning, presence, and attention during ad exposures are consistent with a theory that

Figure 1. Densities of Average Viewer Behaviors



Note. The three panels show densities of the viewer-level average behaviors among the 3,659 viewers.

Figure 2. (Color online) Covariation in Viewer-Level Average Tuning, Presence, and Attention



Note. Each panel presents pairwise comparisons of viewer average (a) tuning and presence, (b) tuning and attention, and (c) presence and attention behaviors during advertising exposures.

older viewers are more likely to avoid ads by changing channels and younger viewers are more likely to avoid ads by leaving the room or diverting their visual attention.

Online appendix Figure A1 graphs mean tuning, presence, and attention by ad slot, based on OTS and exposures. Exposure data show that average tuning gradually rises as ad-averse viewers select out of the break. Average presence also rises uniformly after the first slot. Average attention is nearly constant at 13.5% after the first slot. Together, these findings suggest that passive measurements of viewer presence and attention offer richer information than tuning alone.

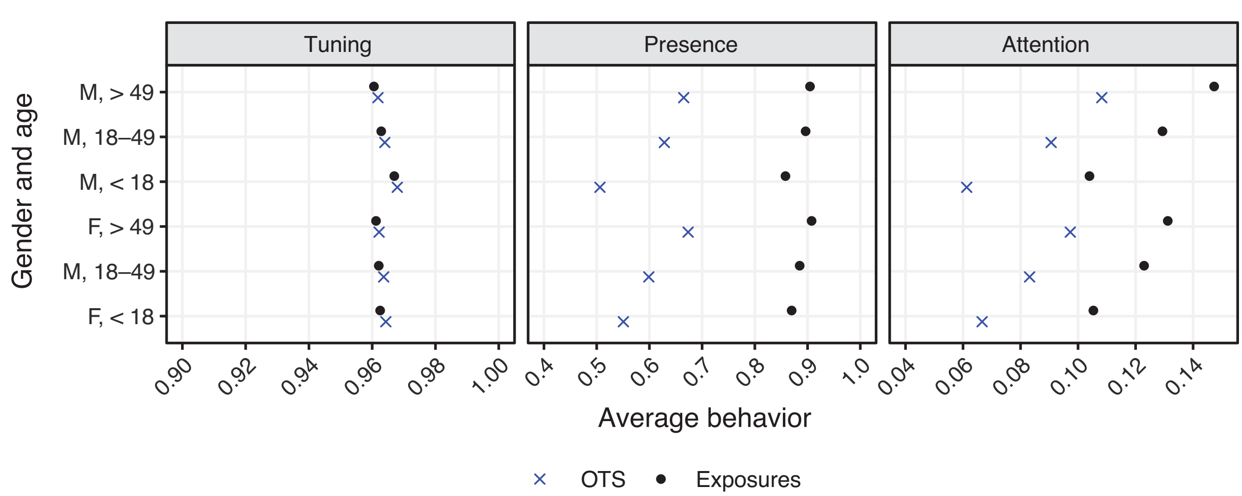
2.2. Sample Selection, Ad Features, and Preliminary Evidence

TVision ad insertion data document the ad environment—network, date, airtime, program, genre, and episode. Ad metadata provide the ad creative name, product name, brand name, product category, and ad duration.[3]

Ad exposures and attention build throughout the day and peak during the evening prime time hours (see online appendix Figure A2). The estimation sample selects ad insertions between 7:00 a.m. and 1:00 a.m. on the four major broadcast networks (ABC, CBS, Fox, and NBC) from July 2016 to June 2017. In total, we observe 4,257,112 exposures of 3,659 unique viewers to 6,650 unique ad creatives in 167 product categories. This sample underpins all estimates reported in Sections 4.1–4.4.[4]

2.2.1. Ad Features. The three most general sets of ad features are ad creative identifiers, brand identifiers, and product category identifiers. An ad creative identifier summarizes all content in a unique ad creative and bounds the behavioral variation ad content could explain. Ad creatives with fewer than 50

Figure 3. Ad Viewing by Viewer Gender and Age



exposures are grouped into a composite, covering 2% of all exposures.

Advertised product categories describe things like beer, cancer drugs, and pick-up trucks. Product categories capture stylistic and thematic similarities across ads, such as humor and good times in beer ads or images of toughness in truck ads. They also reflect regulatory requirements about ad content, such as treated condition or potential side effects in drug ads (Food and Drug Administration2020) .

We follow a long literature on ad content (Resnik and Stern 1977, Anderson et al. 2013, Liaukonyte et al. 2015, Tucker 2015, Anderson et al. 2016, Lee et al. 2018, Tsai and Honka 2021) and supplement ad metadata with three sets of ad features. First, a TV advertising measurement company called iSpot.tv provides an online database of TV ads. We algorithmically downloaded ad videos from iSpot covering 85% of exposures to national ads in the estimation sample (~65% of uniquely labeled ad creatives). We also scraped ad content features from iSpot webpages. For each ad, we observe a tagline identifier, a sentiment score ranging from zero to one based on the positivity of the words in the audio transcript, a promotion identifier, a commercial music identifier, a movie identifier, an engagement rating based on the volume of digital activity related to the ad creative, and a professional actors indicator. iSpot also classifies the mood of each ad as active, emotional, funny, informative, or sexy.

Second, we constructed a set of machine-coded ad content features using machine learning algorithms collated by Schwenzow et al. (2021). We retained measures with precision and recall scores of at least 50%, including number of scene transitions; average colorfulness, saturation, and luminosity; and percentages of ad video seconds showing facial expressions of surprise, happiness, or neutral emotion.[5]

Finally, we used Google Cloud Vision (GCV) to tag recognized images within ad videos. GCV identifies more than 1,000 common image tags in 70 categories, based on a large validation set of human-tagged images and videos.[6] We took two steps to filter out tags likely to be inaccurate or redundant. First, we sought to limit errors in variables by only retaining tags that describe concrete nouns and verbs. Second, we sought to limit collinearity by retaining only those 32 tags for which 50% or more of variance remained unexplained in a regression of the tag on product category, iSpot and machine-coded ad features.

Table A1 in the online appendix summarizes iSpot and machine-coded features, and Table A2 displays GCV image tags and their frequencies in bold. We note three important caveats. First, ad content feature coding is incomplete because no current method can fully characterize video content in interpretable ways. Unobserved features may correlate with coded features and complicate interpretation of feature coding results. Second, classical errors-in-variables issues may bias ad content feature parameter estimates toward zero. Third, ad videos were unavailable for 35% of all creatives covering 15% of sample exposures, so all ad content features implicitly interact video availability with feature measurement.

2.2.2. Linking Ad Viewing to Ad Features. Figure 4 shows how ad viewing changes with ad environment and ad features during ad exposures. Broadcast networks with lower average tuning tend to have higher average attention, a pattern that repeats when comparing prime time to other dayparts. Program genres show some different patterns. For example, ads during football games have the highest tuning and highest attention whereas drama ads have the lowest tuning with moderate attention.

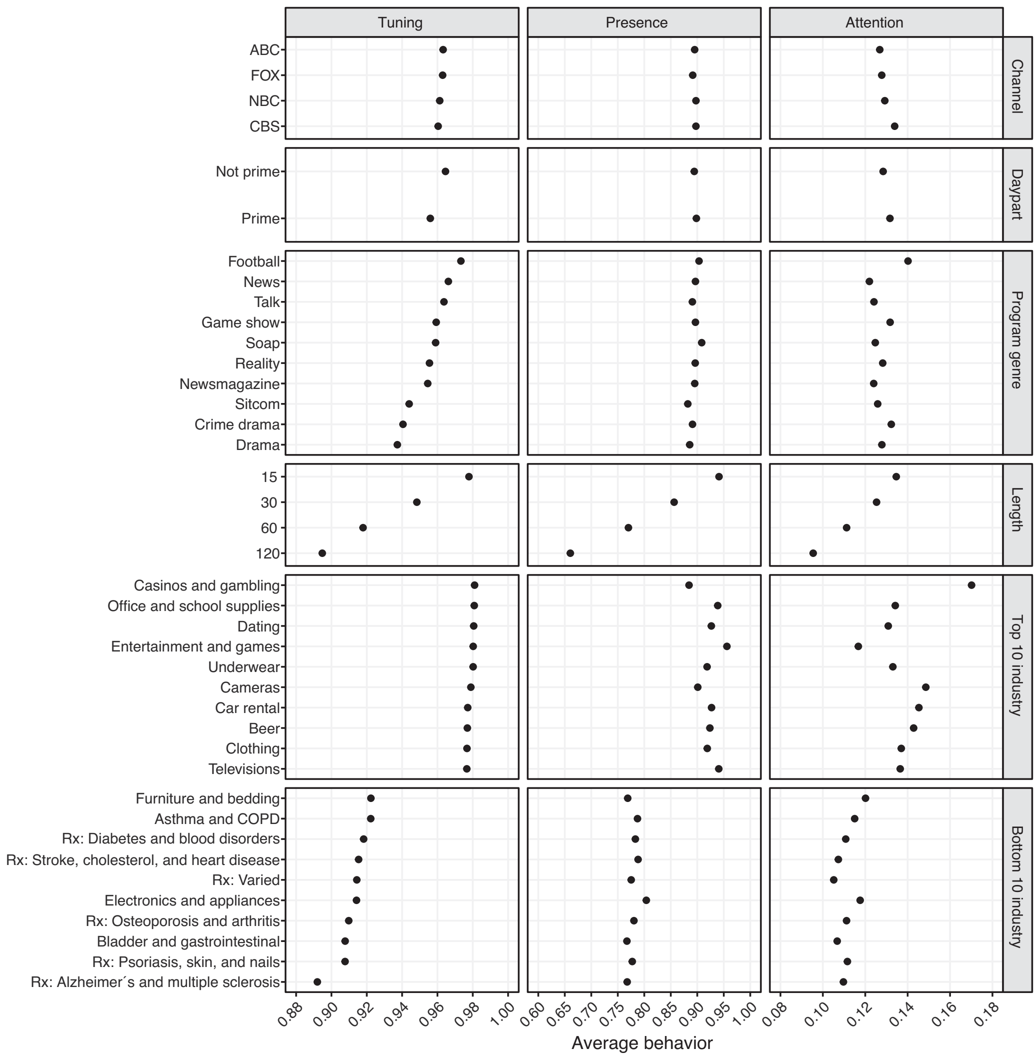
Shorter ads retain more viewers and attention than longer ads on a per-second basis. Comparing 15-second ads to 30-second ads, mean tuning per ad second falls from 98% to 94%, mean presence falls from 94.1% to 85.6%, and mean attention falls from 13.5% to 12.5%.

Advertised product category also correlates with ad viewing. Figure 4 provides mean viewing behaviors for the 10 most-tuned and the 10 least-tuned advertised product categories. The casinos and gambling category is both the most-tuned and most-attended ad category. Entertainment and games ads are highly tuned but attended much less, perhaps because they are more likely to generate second-screening behaviors. Eight of the 10 least-tuned ad categories are for prescription drugs, and they all receive less attention than the remaining two least-tuned product categories.

Table A3 in the online appendix presents variance decompositions of ad viewing behaviors on individual sets of viewer, break, and ad features. Viewer identifiers are the best predictors of presence and attention, explaining 53 times more variation in attention than the traditional targeting variables of age and gender combined. This finding congrues with prior research quantifying the profitability of individually targeted advertising (Deng and Mela 2018). Ad environment variables also correlate with viewing behaviors, including slot within the break, network, and program genre.

The ad features correlate weakly with presence but explain less variation in tuning and remarkably little variation in attention. One of the strongest correlates is ad duration, explaining 1.6% of tuning and 4.6% of presence but just 0.1% of attention. Another is ad category, which explains 0.5% of tuning, 1.0% of presence, and 0.1% of attention. In summary, the variance

Figure 4. Viewing Behaviors During Ad Exposures by Break and Ad Characteristics



Note. This figure presents average tuning, presence, and attention behaviors across channels, dayparts, program genres, ad lengths, and top and bottom ad categories (as ranked by average tuning).

decompositions presage difficulty in detecting effects of ad content on advertising attention.

3. Empirical Framework

We describe the model, causal identification, and results interpretation.

3.1. Model Specification

We develop an empirical model in the causal effects paradigm described by Chintagunta and Nair (2011). The model explains tuning, presence, and attention behaviors as functions of ad features, slot and time-within-break features, and viewer-break interaction effects. Subscript b indexes ad breaks, each of which is a set of consecutive ads inserted into a specific network-program-date-time combination. Each ad slot within a break is indexed with s, so every (b, s) combination identifies a unique insertion of the particular ad creative that was aired in slot s of break b.

Let y^j_{ibs} be viewing behavior j for viewer i exposed to the sth advertisement in ad break b.[7] Ad viewing behavior is modeled as follows:

$$y^j_{ibs} = x'_{bs}\boldsymbol{\beta}^j + g(1_s, l_{bs}, t_{bs}; \Theta^j) + \delta^j_{ib} + \varepsilon^j_{ibs}, \qquad (1)$$

where x_{bs} is a vector of ad features, such as ad creative fixed effects, or product category fixed effects and ad content features; and $\boldsymbol{\beta}^j$ represents how ad characteristics change mean viewing behaviors.[8]

The function $g(1_s, l_{bs}, t_{bs}; \Theta^j)$ estimates average changes in viewing behaviors during commercial breaks. The slot-specific indicator variables 1_s capture typical changes in viewing behaviors across ad slots. Ads in the sample range from 15 to 120 seconds, so it is also important for g to accommodate differences in advertisement durations, denoted l_{bs}, as well as the total time elapsed since the beginning of the break, t_{bs}.

The parameter δ_{ib}^{j} captures heterogeneity across viewers, breaks, and viewer breaks. It inherently nests. (i) viewer-specific effects including viewer habits or viewing environment idiosyncrasies; (ii) break-specific effects including time, program, or network shocks, for example, the program in which the break occurs, how much time has passed since the last break, the season of the year, or the time of day; and (iii) viewer-break interaction effects, such as how engaged viewer i is with the program or whether the viewer is watching the break during time-shifted programming. The flexibility of δ_{ib}^{j} comes from its high dimensionality given that there are 994,186 (i, b) combinations in the estimation sample. We estimate $\delta_{i,b}$ parameters using the method of alternating projections (Guimaraes and Portugal 2010).[9]

The error term, ε_{ibs}^{j}, captures any remaining omitted factors such as measurement error.

3.2. Causal Identification: Theory and Evidence

A small but growing literature has recently established that advertising endogeneity problems can be unusually severe. Lewis and Rao (2015) showed that small model specification errors can overwhelm treatment effects of digital banner ads on sales. Gordon et al. (2019) found that observational methods failed to recover experimental treatment estimates of Facebook ads on sales, even in huge samples with numerous covariates. Shapiro et al. (2021) found that careful endogeneity controls estimated smaller effects of TV ads on packaged good sales than correlational approaches.

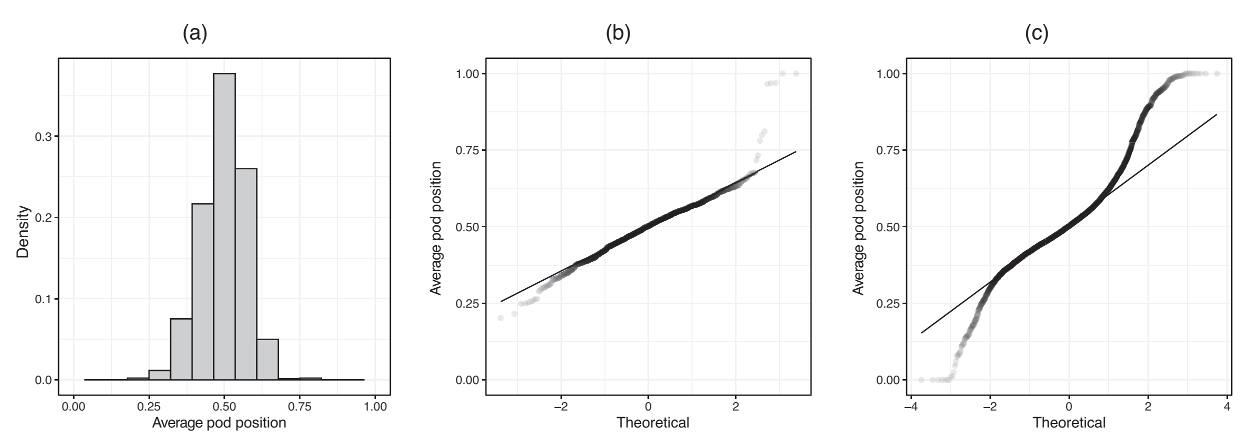
The ideal experiment to identify ad content effects on viewing behaviors would randomize ads across audiences, brands, ad breaks, and slots. However, we know that advertisers and viewers both self-select into commercial breaks (Tuchman et al. 2018). Therefore, we assume x_{bs} correlates with δ_{bs}^{j} in the causal models and estimate the δ_{bs}^{j} parameters. For comparison, we also report results of descriptive models in which the δ_{bs}^{j} parameters are treated as unobservables.

We then rely on broadcast TV networks' quasi-random ordering of ads within breaks, which implies x_{bs} is uncorrelated with ε_{ibs}^{j}. The television industry has long known or assumed that viewing behaviors change across ad slots within the commercial break, as confirmed in Figure A1 in the online appendix. However, broadcast networks do not sell specific ad slots to advertisers. Advertisers purchase ad insertions based on networks, dates, and quarter-hours, typically months in advance and without guarantees of what program the ad will be inserted into. The exclusion of ad slots from standard ad contractual terms can be explained by observing that Nielsen audience estimates do not vary meaningfully between consecutive ad slots, likely due to the relative imprecision of People Meter presence measurements. Instead, standard TV ad sales contracts promise to rotate ads across slots on an equitable basis across commercial breaks (Mandese 2004).

Quasi-random ordering of TV ads within commercial breaks is verifiable. If networks assign ads to slots using independent random draws, then the distribution of ad creatives' average slots should be normal, by the law of large numbers. To check, we define each ad insertion's position within its break as $\frac{s-1}{S_b-1}$, where S_b is the number of slots in break b. Thus, every ad position lies in [0, 1] and the measure is comparable across various ad and break durations.

Figure 5(a), plots the empirical distribution of average ad positions for the 1,384 advertised products with at least 50 ad insertions on broadcast networks.[10] The distribution appears approximately normal. Panel (b) compares the empirical distribution of average ad positions to

Figure 5. Randomization Check



quantiles of a normal distribution with the same mean and variance. There is a remarkably close correspondence. All eight of the largest positive outliers are ads for sports programs that were most likely house ads run by program producers (e.g., NFL Online, USGA Organization, FedEx Sponsored Event). Overall, ad positions are verifiably consistent with networks' contractual promises of quasi-random ad ordering.

Quasi-random ordering does carry an important caveat. Some cable networks price ads by slot. In fact, average cable network ad slots depart meaningfully from quasi-random placement, as shown in Figure 5(c). Therefore we excluded cable networks from the sample. Quasi-random ordering is also unlikely in addressable TV or other programmatic ad sales contexts.

Numerous papers rely on quasi-random ordering of TV ads within breaks to identify causal TV ad effects.[11] However, to the best of our knowledge, no prior study has empirically confirmed quasi-random assignments of ads to slots, so the confirmation may be a contribution to methodologically similar studies.

3.3. Interpretation of Effects

Consumers often use ad blockers or digital video recorders to avoid ad exposures, but avoidance behavior is seldom observable in advertising data. Therefore, most advertising studies estimate intent-to-treat (ITT) effects (Tuchman et al. 2018, Gordon et al. 2021). TVision data are unusual in that they enable direct measurements of ad treatments (i.e., exposures), enabling us to distinguish treatment effects from ITT effects.

We interpret the ad content effects on viewing behaviors as local average treatment effects (LATE) (Imbens and Angrist 1994). LATE, by definition, conditions on all forms of selection, including advertiser and viewer selection into breaks. LATE estimates quantify how targeted ads changed viewing behaviors within targeted contexts, and therefore can inform advertiser evaluations of ad effects within targeted contexts.

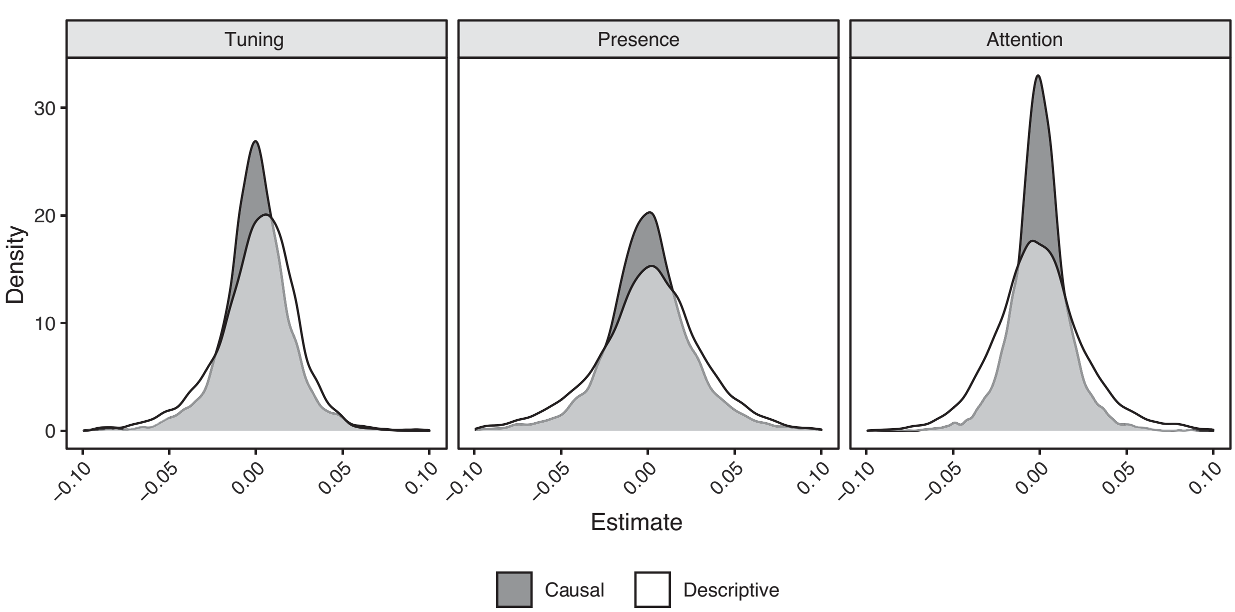
LATE estimates may not extrapolate to untreated contexts, motivating the question of whether average treatment effects (ATE) are estimable. If all potential contexts were treated, then LATE and ATE are equal. However, that is unlikely, because viewers and advertisers each select into breaks, and as a result advertised categories can co-occur within breaks.[12]

We sought to quantify the extent of advertiser targeting in the data. The average viewer is exposed to ads from 39% of the 167 ad categories, and the exposure-weighted average is 70%. The average ad category co-occurs within the same break as 76% of the 166 other ad categories, and the exposure-weighted average is 93%. Figure A3 in the online appendix illustrates the distributions of viewer category exposures and category co-occurrences within breaks whereas Figure A4 plots coverage of ad categories against each viewer sorted by exposures. Overall, we conclude that broadcast TV ad targeting seems limited, consistent with its perceived role as a mass medium.

4. Findings

We present ad creative and category effects; duration, slot, and time effects; ad feature effects; robustness

Figure 6. Distributions of Ad Creative Estimates



Note. Distributions are demeaned to aide comparisons across models and outcomes.

checks; and a deeper case study of drug category ad effects.

4.1. Ad Creative Results

Figure 6 depicts six distributions of ad creative effects: one each for the descriptive and causal models, within each of the tuning, presence, and attention regressions. Each distribution characterizes 6,650 parameter estimates. We demean the distributions to focus on their shapes.

All six distributions are unimodal and nearly symmetric. The causal effect distributions vary less than their descriptive counterparts, with the greatest compression observed in the attention estimates. Specifically, the standard deviation of the causal tuning distribution is 12% smaller than in the descriptive tuning distribution, 20% smaller for presence, and 37% smaller for attention.

The greater variation in the descriptive distributions shows that ad effects on viewing behaviors covary with factors that predict ad assignments to highly viewed breaks, such as viewer factors, break factors, and viewer-break factors. The only difference in the models that generate the different results is whether the δ_{ib}^{j} parameters are estimated jointly with the ad creative effects or treated as unobservables as part of the error term.

Still, despite the compression within the causal effect distributions, the tails of those distributions contain some surprisingly large ad creative point estimates. For example, 5% of the point estimates in the causal tuning distribution exceed 0.039 in absolute value, more than the difference between tuning's average and its upper bound (0.963 and 1.0, respectively). 5% of the presence point estimates exceed 0.048 in absolute value, and 5% of the attention point estimates exceed 0.030 in absolute value, both of which are surprisingly large compared with sample averages (e.g., 0.129 for attention), especially when considering that the regression separately accounts for slot effects and time-into-break effects.

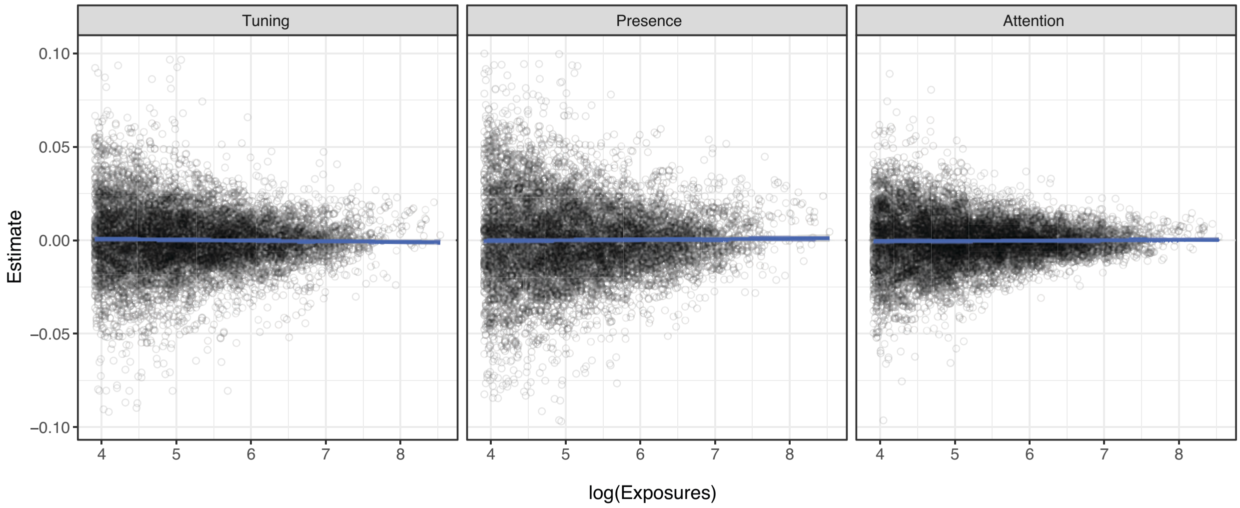
We sought to better understand how individual ad creative estimates relate to sample sizes. Figure 7 turns the causal distributions on their side, showing how creative estimates covary with log(exposures). The x axis runs from e^4 to e^9, indicating that sample ad creatives vary 100-fold in total exposures, from 50 to 5,076. The trend lines show that the creative estimates are nearly uncorrelated with the number of exposures. The outlying estimates are all infrequent ads; the most frequently viewed ad creative estimates are more concentrated around zero.

It is possible to bootstrap ad creative standard errors, but we prefer not to risk interpreting noise. We also investigated replacing the ad creative fixed effects with the 1,504 brand fixed effects, but again found a pattern quite similar to Figures 6 and 7: outlier estimates tended to be brands with limited ad exposures and thus indistinguishable from noise. We focus instead on a model that shrinks the ad creative effects toward product category identifiers and ad content features.

4.2. Ad Category Results

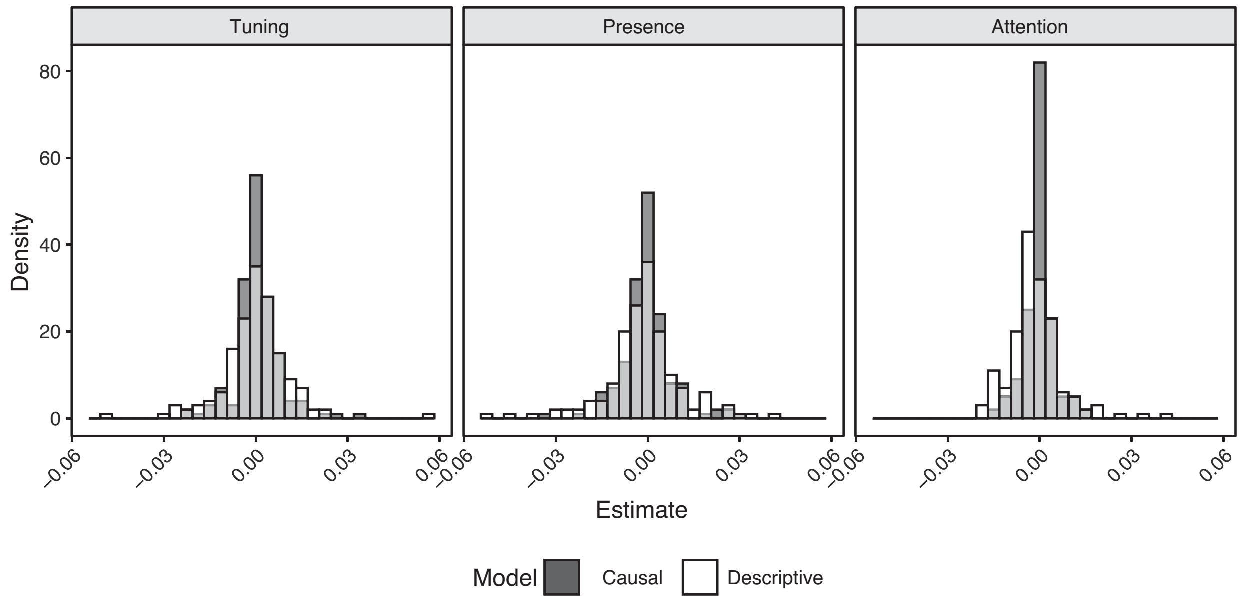
Figure 8 depicts the causal and descriptive distributions of product category effects. These distributions are unimodal with the causal distributions again being

Figure 7. (Color online) Ad Creative Estimates by Sample Size



Notes. The three panels show scatterplots of causal ad creative estimates versus the log number of exposures for each ad creative. Distributions are demeaned to aide comparisons across models and outcomes. The trend line shows linear fit.

Figure 8. Distributions of Ad Category Estimates



Notes. The three panels show distributions of 167 ad category estimates on tuning, presence, and attention from the descriptive and causal models. Distributions are demeaned to aide comparisons across models and outcomes.

tighter than their descriptive counterparts. However, the degree of compression is similar across the three viewing behaviors. We observe 22%, 26%, and 21% decreases in standard deviations between the descriptive and causal models for tuning, presence, and attention, respectively ($SD^T_{Desc} = 0.0010$, $SD^T_{Caus} = 0.008$; $SD^P_{Desc} = 0.013$, $SD^P_{Caus} = 0.009$; $SD^A_{Desc} = 0.009$, $SD^A_{Caus} = 0.007$). Ad category effect distributions are tighter than ad creative effect distributions: 5% of the point estimates in the causal tuning distribution exceed 0.015 in absolute value, 5% of the presence point estimates exceed 0.017 in absolute value, and five of the attention point estimates exceed 0.011 in absolute value.

Table 1 shows evidence that product category estimates contain reliable signals. A 5% error rate predicts 8.35 false positives among the 167 category effects in any of the six regressions due to random chance alone. The causal model results indicate 32 significant category effects on tuning, 20 significant effects on presence, but only eight significant effects on attention. Therefore, we report but do not really interpret the category effects in the attention regression.

Figure 9 shows that ad category parameter significance is not driven by low-powered outliers. The majority of significant results do not occur among the lowest-powered coefficients. The tuning and presence panels show that many significant positive and significant negative category causal effects are well powered.

Figure 10 unpacks the results presented in Figure 9, highlighting the 40 highest and lowest ad category fixed effects as ranked by tuning estimates. Many of the largest category effects relate to recreation, including hunting and fishing; casinos and gambling; wine, alcohol, and e-cigarettes; dating; sports; movies; and airlines, whose television ad content promotes leisure travel. Many of the most negative category effects relate to prescription drugs, including drug categories treating cancer; depression, bipolar, and insomnia; Alzheimer's and multiple sclerosis; psoriasis, skin, and nails; osteoporosis and arthritis; varied conditions; bladder and gastrointestinal; and stroke, cholesterol, and heart disease.

The presence regression estimates mostly align with the tuning results but exhibit larger standard errors. The categories that reliably retain viewer presence include wine, alcohol, and e-cigarettes; underwear; car rental; sports; clothing; speakers and headphones; movies; legal services; and shoes and socks. On the other end, six of the seven largest significant negative findings again feature prescription drugs.

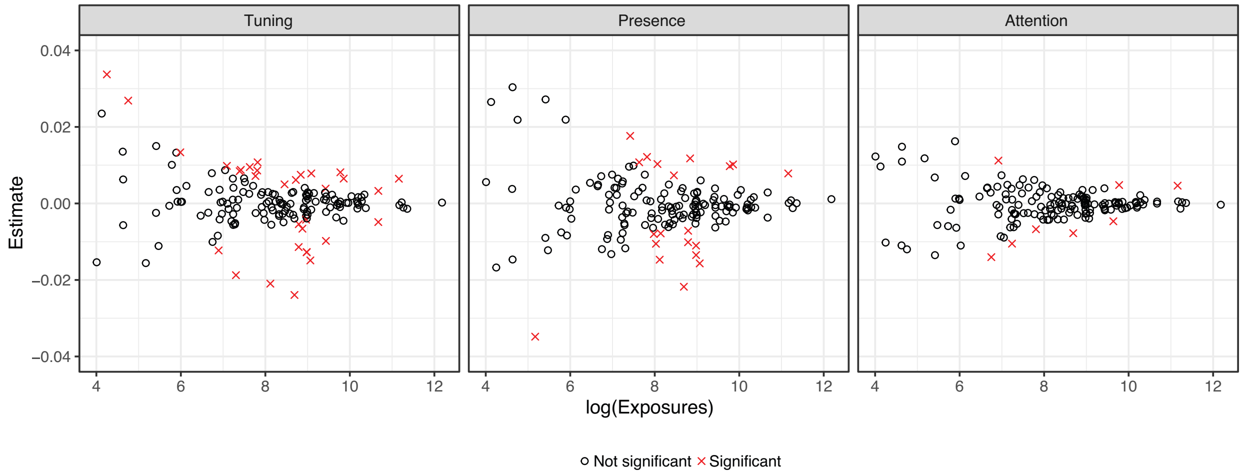
Sixty-six of the 80 category effects on attention have confidence intervals that lie entirely between −2% and 2% of ad seconds.[13] Attention requires both tuning

Table 1. Counts of Significant Ad Category Estimates

Model	Tuning	Presence	Attention
Descriptive	68	45	33
Causal	32	20	8

Note. Table entries count how many of the 167 ad category estimates are statistically significantly different from the average ad at the 95% confidence level.

Figure 9. Ad Category Estimates by Sample Size



Notes. The three panels show scatterplots of causal ad category estimates versus the logged number of exposures. Distributions are demeaned to aide comparisons across models and outcomes. A red mark indicates that an estimate's 95% confidence interval does not include zero.

and presence, by definition, so we expected attention results to resemble tuning and presence results. It is possible that ingrained habits drive viewer attention more than on-screen ad content. Another possibility is attention measurement error may lead to larger standard errors than other regressions. A third explanation could be that ad content is polarizing: if some content reliably attracts attention from viewers interested in the product market, it may simultaneously lead uninterested viewers to divert their attention, change channels, or leave the room.

The ad creative and category results highlight the importance of controlling for factors that predict ad assignment to highly viewed breaks. Namely, ad estimates are inflated if viewer-break effects are not explicitly taken into account. Whereas most ad category effects are not large, there are notable significant effects—viewers tune away during prescription drug ads one to three percentage points more frequently than during a typical ad, which is 25%–75% more than the baseline average tune-out rate of about 4%.

4.3. Duration, Slot, and Time Effects

Ad duration, slot, and time effects help characterize ad viewing dynamics within ad breaks. Table 2 presents ad duration, slot, and time-elapsed effect estimates. Figure 11 graphs how ad durations change viewing behaviors. Thirty-second ads reduce tuning by an absolute 2.7%, presence by 5.6%, and attention by 0.7%. Sixty-second ad duration effects are approximately double the absolute 30-second duration effects, per ad second. Sixty-second ads reduce tuning by an absolute 5.5%, presence by 10.9%, and attention by 1.5%.

These results suggest that longer ads perform relatively worse (on a per-second basis) at keeping viewers tuned-in, present, and attentive to ads. This may explain the shift in recent years toward shorter duration TV advertisements (Friedman 2017).

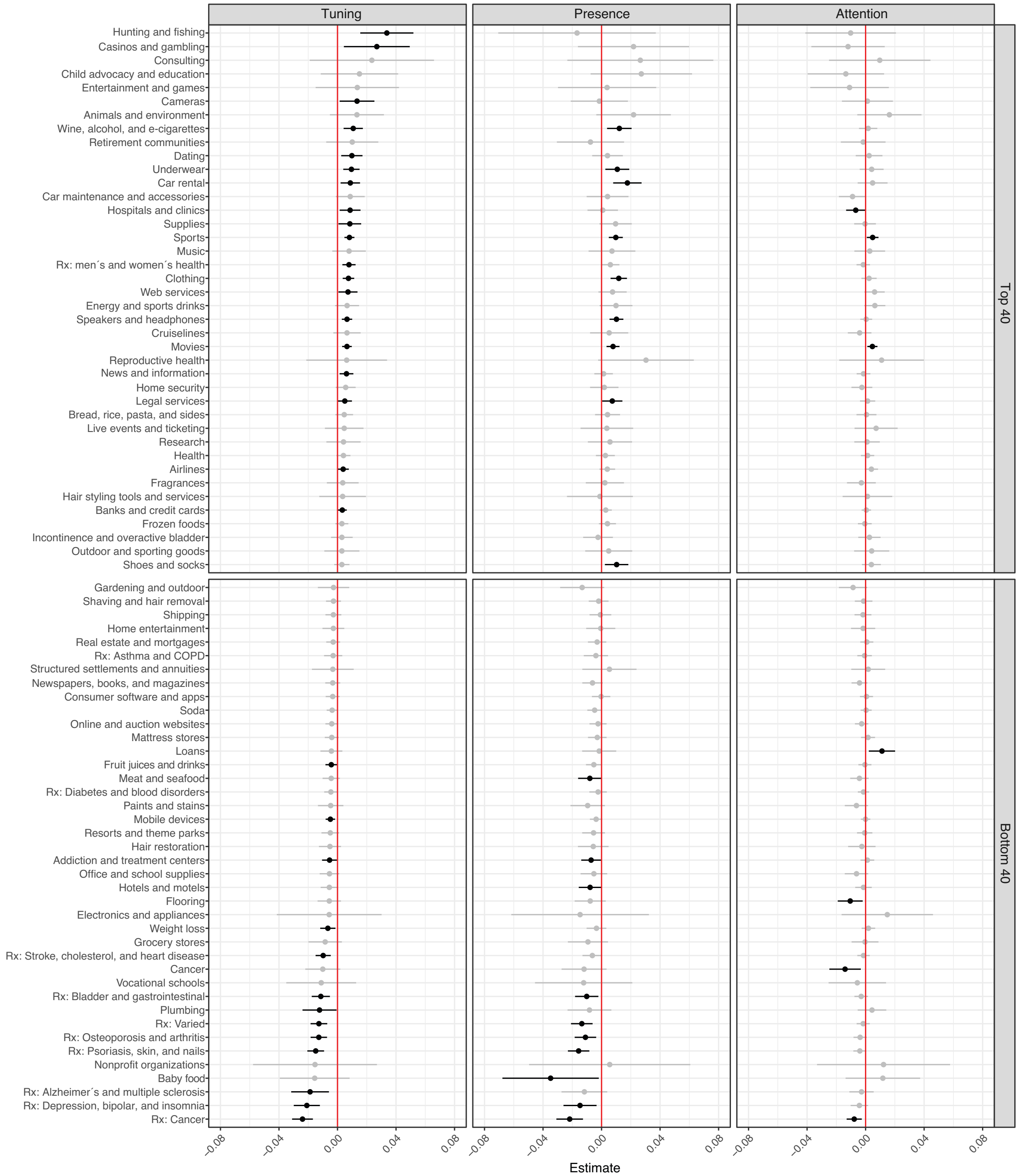
Next, we look at slot and time-elapsed effects in Table 2. The modal ad break contains seven slots and the modal ad duration is 30 seconds. Figure 12 graphs total effects of ad slot, ad duration, and time-elapsed effects on viewing behaviors for a hypothetical break consisting of seven 30-second ads. Standard errors of the combined effects are calculated by bootstrapping out of the joint asymptotic distribution of the parameter estimates, including off-diagonal terms. Confidence intervals widen throughout the break because audience-remaining calculations in later slots depend on earlier slots' parameter estimates.

Tuning decreases across slots with changes driven primarily by the time-elapsed variables. Presence shows a similar absolute decrease but is significantly impacted by both slot effects and time-elapsed variables, as shown in Table 2. Attention decreases in the first two slots, but confidence intervals overlap from positions three to seven, and time-elapsed variables are not significant predictors. Therefore, although the audience decreases throughout the break, attention does not always decrease after the third slot. These patterns suggest that viewers who do not tune away or leave the room early in the break are, on average, more attentive to ads than viewers who leave during the first few slots of the break.

4.4. Ad Feature Results

We measured ad content because ad viewing behaviors may respond to stimuli displayed on television screens. However, we interpret the following ad feature

Figure 10. (Color online) Top and Bottom 40 Ad Category Causal Effects

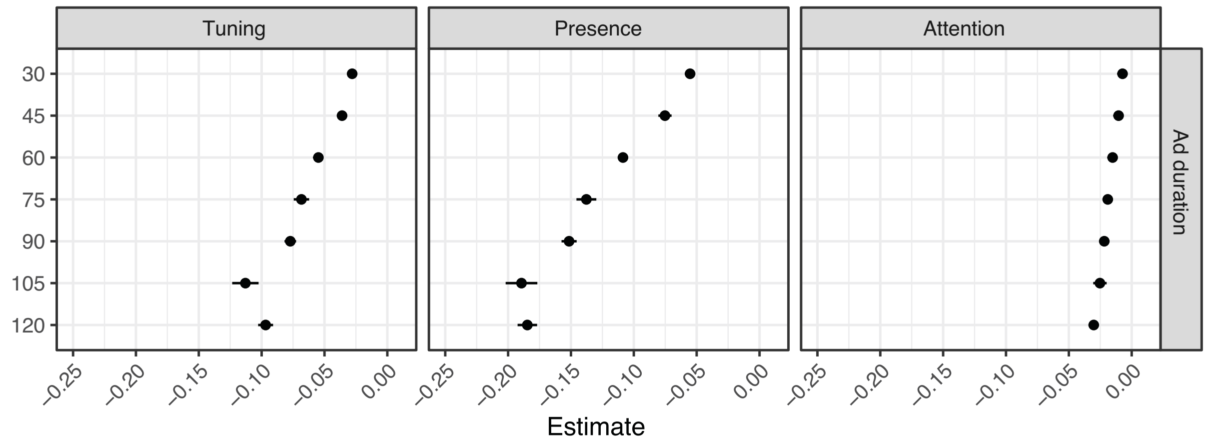


Notes. Whiskers represent 95% confidence intervals. Bolded point estimates and whiskers indicate 95% confidence intervals that do not include zero.

results with caution, given the caveats about unobserved features, feature measurement error, and feature availability. Still, these effects are valuable controls in that they help separate ad category effects from common features included in those categories. Additionally, whereas individual coefficients may be difficult to interpret, there may be interesting patterns in ad feature estimates across viewing behaviors.

Figure 13 presents iSpot and machine-coded ad feature causal effects on tuning, presence, and attention.

Figure 11. Ad Duration Effects



Notes. Ad durations are measured in seconds. Whiskers represent 95% confidence intervals. Every 95% confidence interval in the graph does not include zero.

As before, tuning and presence results are more precisely estimated than attention, but all of the effects are small on an absolute basis. Sales-related content such as taglines and promotions reduce tuning and presence, similar to findings in Teixeira et al. (2010). Surprisingly, a higher sentiment score reduces tuning, and professional actors reduce both tuning and presence. It seems plausible that brands may try to make otherwise unattractive ad messages more palatable by using professional actors or more positive scripts.

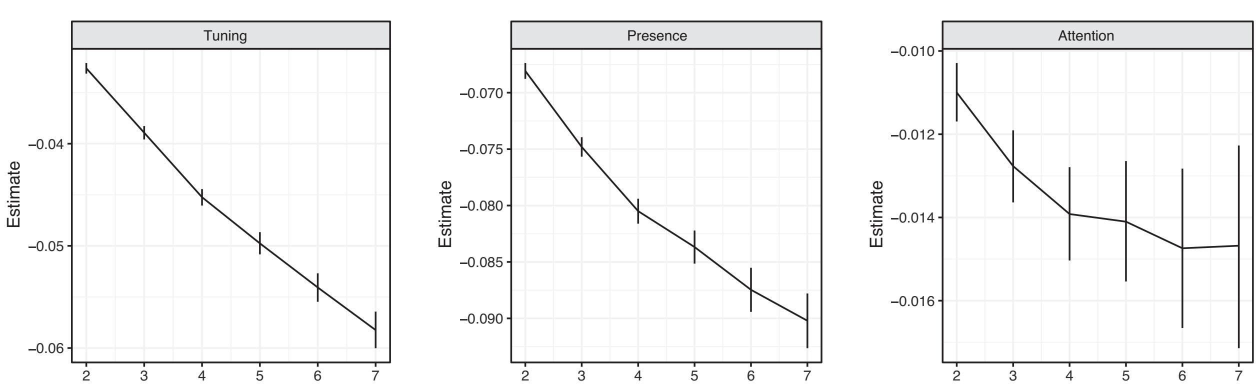
Another surprise is that the movie dummy reduces tuning and presence, given that the movies product category effect increases ad viewing. We investigated this more carefully by examining the overlap between the movies category dummy and the iSpot movie classifier. The iSpot variable indicates both theatrical movie trailers and the presence of theatrical movie brands in cobranded advertisements, such as for packaged goods, cars, fast food, and retail chains. If we drop the movies category dummy, then the iSpot movie effect becomes positive and significant. Therefore, it appears that movie ads increase viewing, but movie cobranding in nonmovie ads reduces tuning and presence.

The engagement variable measures ad traffic across iSpot's video, social, and search channels, and increases ad viewing behaviors. Ads classified by iSpot as having a "sexy" mood reduce tuning, but other mood variable effects are estimated imprecisely.

Two of the machine-coded features have significant effects. The number of scenes within an ad reduces tuning and presence, as does the duration of neutral facial expressions shown on screen. Other machine-coded features generally have point estimates near zero.

Figure 14 presents the effects of Google Cloud Vision features on viewing behaviors. The confidence intervals are much smaller than in Figure 13 and the

Figure 12. Position and Time-Elapsed Effects in a Modal Break



Note. Whiskers represent 95% confidence intervals.

Table 2. Slot, Duration, and Time-Elapsed Parameter Estimates

Variable	*Tuning*	*Presence*	*Attention*
Pod position			
2	0.00049	−0.00816***	−0.00352***
	(0.00036)	(0.00048)	(0.00040)
3	−0.00068	−0.01024***	−0.00509***
	(0.00056)	(0.00076)	(0.00067)
4	−0.00185*	−0.01123***	−0.00603***
	(0.00075)	(0.00101)	(0.00090)
5	−0.00124	−0.00976***	−0.00602***
	(0.00093)	(0.00125)	(0.00110)
6	−0.00042	−0.00891***	−0.00643***
	(0.00108)	(0.00145)	(0.00129)
7	0.00053	−0.00700***	−0.00617***
	(0.00122)	(0.00164)	(0.00146)
8	0.00194	−0.00632***	−0.00658***
	(0.00135)	(0.00182)	(0.00162)
9	0.00128	−0.00675***	−0.00763***
	(0.00148)	(0.00199)	(0.00178)
10	0.00163	−0.00586**	−0.00689***
	(0.00161)	(0.00217)	(0.00195)
11	0.00125	−0.00609*	−0.00827***
	(0.00175)	(0.00236)	(0.00214)
12	0.00407*	−0.00356	−0.00783***
	(0.00190)	(0.00258)	(0.00235)
13	0.00363	−0.00574*	−0.00845**
	(0.00208)	(0.00283)	(0.00261)
14	0.00223	−0.00916**	−0.00974***
	(0.00231)	(0.00315)	(0.00291)
15	−0.00025	−0.00844*	−0.00908**
	(0.00262)	(0.00356)	(0.00336)
16	−0.00640*	−0.01139**	−0.01059**
	(0.00313)	(0.00418)	(0.00383)
17	−0.01824***	−0.02730***	−0.00937*
	(0.00406)	(0.00530)	(0.00457)
18	−0.04495***	−0.05422***	−0.01302*
	(0.00622)	(0.00774)	(0.00622)
Ad duration			
30	−0.02798***	−0.05526***	−0.00730***
	(0.00018)	(0.00025)	(0.00019)
45	−0.03603***	−0.07523***	−0.01044***
	(0.00197)	(0.00268)	(0.00163)
60	−0.05485***	−0.10864***	−0.01517***
	(0.00070)	(0.00094)	(0.00056)
75	−0.06833***	−0.13772***	−0.01904***
	(0.00313)	(0.00403)	(0.00198)
90	−0.07718***	−0.15149***	−0.02182***
	(0.00235)	(0.00307)	(0.00155)
105	−0.11291***	−0.18936***	−0.02527***
	(0.00535)	(0.00643)	(0.00272)
120	−0.09683***	−0.18470***	−0.03020***
	(0.00303)	(0.00396)	(0.00197)
Time elapsed	−1.749e-04***	−1.624e-04***	−7.167e-06
	(1.216e-05)	(1.647e-05)	(1.424e-05)
Time elapsed^2	1.330e-07***	2.339e-07***	1.601e-08
	(3.267e-08)	(4.625e-08)	(4.339e-08)
R^2	0.366	0.571	0.711
N	4,257,112	4,257,112	4,257,112

Note. Standard errors are in parentheses.
$^*p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

large majority of feature labels do not have significant effects. Further, those few features that do have significant effects resist easy interpretation. For example, one might have predicted that infant or party might have increased viewing, but infant is near zero and party is negative. We again recall the caveat that labeled features may correlate with important unlabeled features, such as when brands pair less attractive selling messages with more attention-grabbing stimuli.

4.5. Robustness Checks

We view the ad category results as the most interesting set of causal effects, so we investigate how stable they are to alternate model specifications. Figure A5 in the online appendix shows that ad category findings are nearly identical when ad content features are removed from the model. Remarkably, none of the 480 depicted category point estimates falls outside the other model's confidence interval, suggesting that the category findings may be highly reliable.

Tuning is a device-level property, but viewing often occurs among groups of viewers. Figure A6 in the online appendix shows that ad category effects on tuning are generally robust to exclusion of multiple-viewer viewing sessions.

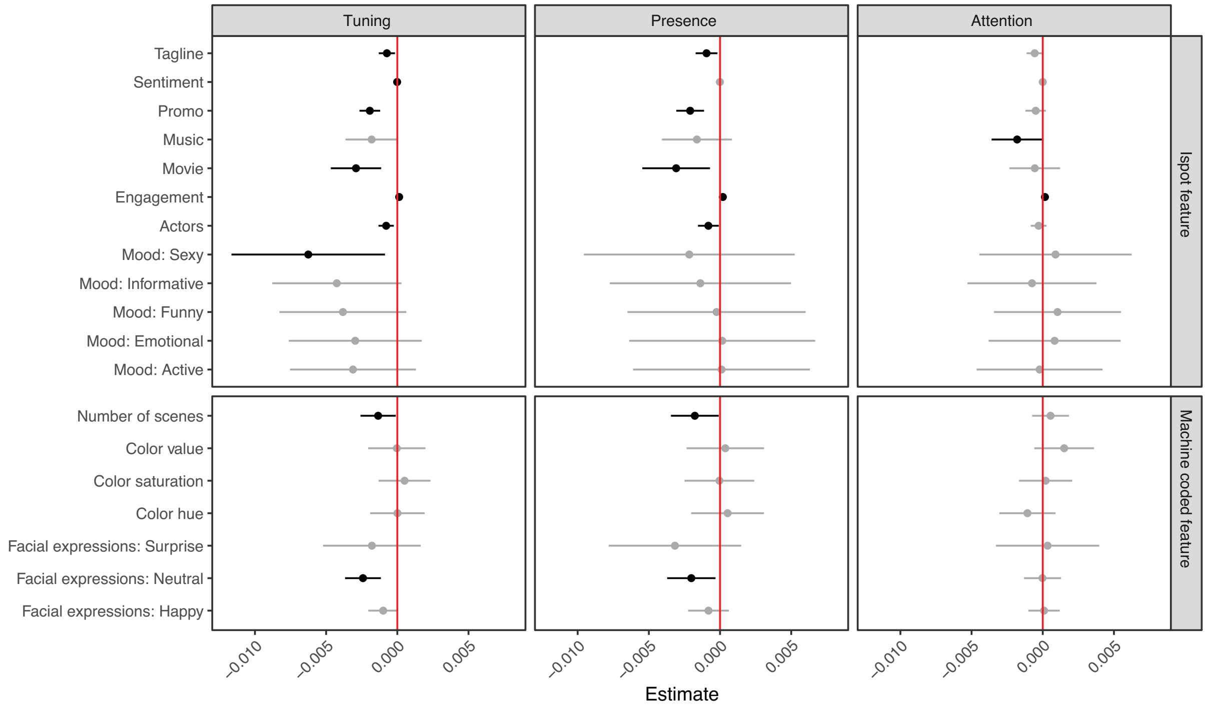
Finally, we investigated the ability of the data to estimate heterogeneous ad effects. Figure A7 in the online appendix shows that ad category confidence intervals overlap when the sample is partitioned between demographic groups.

4.6. Case Study: Prescription Drugs

Pharmaceutical drug advertisements tend to cause viewers to tune away and leave the room. Moreover, the most negative effects relate to serious conditions such as cancer, depression, and Alzheimer's. We quantify how drug category attributes relate to ad effects on viewing behaviors.

4.6.1. Background. Previous research has found that pharmaceutical advertising tends to increase drug category consumption (Iizuka and Jin 2005, 2007; Shapiro 2018, 2022; Sinkinson and Starc 2018). If drug advertising enhances general welfare, then we should care about factors that may influence drug ad pricing. Digital ad sellers Facebook and Google typically include some element of consumer acceptance or rejection of ads in their advertising pricing algorithms, as earlier ads affect attention paid to subsequent ads. Broadcast television networks do not publish ad pricing algorithms, but they too may use audience reactions to price ads. Here, we seek

Figure 13. (Color online) iSpot and Machine-Coded Ad Feature Causal Effects



Notes. Whiskers represent 95% confidence intervals. Bolded point estimates and whiskers indicate 95% confidence intervals that do not include zero.

to quantify what drug category factors correlate with larger or smaller audience losses during drug ads.

Viewers may avoid pharmaceutical ads for two distinct but related reasons. First, drug ads may present viewers with unpleasant reminders of prevalent adverse health conditions.[14] Second, advertised drugs may present viewers with unpleasant reminders of particularly severe health conditions.

4.6.2. Data. We collected objective measures of treated condition prevalence and severity for each pharmaceutical category. Prevalence is measured as the case rate, or the percentage of U.S. residents who experience the disease or condition in a year. Severity is measured in disability-adjusted life years (DALYs), which estimates years of life lost due to premature death and years of healthy life lost due to poor health or disability, reflecting both mortality and morbidity in a single measure. Prevalence and severity measures come from U.S. data compiled by the Institute for Health Metrics and Evaluation's Global Burden of Disease studies.

Table A4 in the online appendix presents the category prevalence and severity data. The two measures correlate at −0.2 at the category level. A few categories have both high prevalence and high severity (e.g., stroke, cholesterol, and heart disease; depression, bipolar, and insomnia). Most other categories have either high prevalence or high severity.

4.6.3. Model and Results. We run a second-stage regression of pharmaceutical category effect estimates from Equation (1) on disease prevalence and severity using Equation (2).

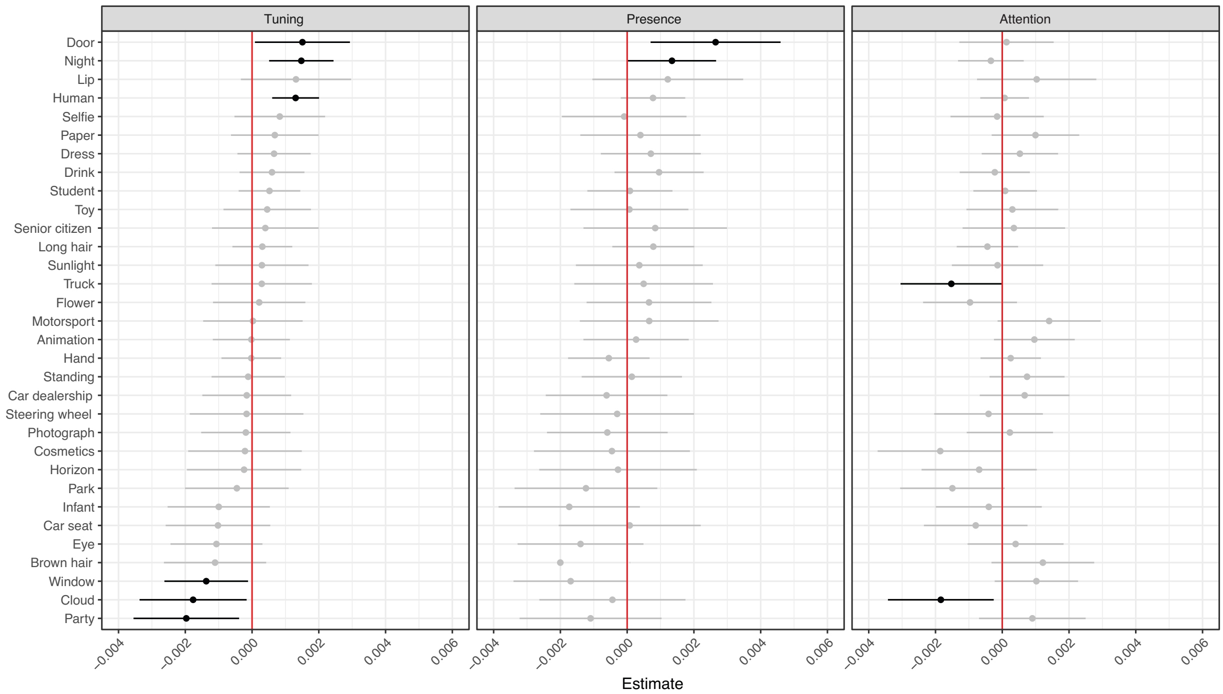
$$\hat{\beta}_k^j = \alpha^j + \gamma_1^j \mathrm{PREV}_k + \gamma_2^j \mathrm{SEV}_k + \varepsilon_k^j, \qquad (2)$$

where $\hat{\beta}_k^j$ is the causal effect estimate of prescription category k ads on viewing behavior j (tuning, presence, or attention); and PREV_k and SEV_k are the prevalence and severity of drug category k.

We account for first-stage estimation error using the estimated asymptotic joint distributions of the point estimates. Specifically, we estimate Equation (2) using generalized least squares (GLS) with the estimated variance-covariance matrix, $\hat{\Omega}_{Rx}^j$, where $\hat{\Omega}_{Rx}^j$ is the relevant subset of the full variance-covariance matrix of the parameters estimated in the first-stage regression, covering only the pharmaceutical category effects, including the off-diagonal terms. Intuitively, estimating Equation (2) via GLS gives more weight to the more precise drug category estimates.

Table 3 presents the estimation results from Equation (2). The table also shows restricted models

Figure 14. (Color online) Google Cloud Vision Ad Feature Causal Effects



Notes. This figure presents Google Cloud Vision (GCV) feature estimates, ranked by tuning estimates. Features are tags created by GCV based on ad creative videos. Whiskers represent 95% confidence intervals. Bolded point estimates and whiskers indicate 95% confidence intervals that do not include zero.

that contain each predictor individually. We view the results in column (3) as the most informative specification. We interpret the results as descriptive as drug firms choose ads strategically and no instruments are available for unobserved category attributes.

Drug category prevalence and severity are both associated with more negative category effects on TV ad tuning and presence. The point estimates show increasing drug category severity by one million DALYs is associated with a 0.07% decrease in tuning and a 0.05% decrease in presence. Increasing drug category prevalence by 1% of the population correlates with a 0.05% decrease in tuning and a 0.5% decrease in presence.

5. Can Viewing Metrics Predict Ad Response?

Next, we investigate whether new ad viewing metrics can predict audience response to ads. To do this, we require an ad response measure that is publicly available, reliably measured, temporally disaggregated, relevant, and comparable across brands. We focus on online brand search, which has been studied extensively in relation to TV ads (Zigmond and Stipp 2010, Chandrasekaran et al. 2018, Du et al. 2019, Lambrecht et al. 2021, Liaukonyte and Žaldokas 2021) and found to predict changes in brand attitudes (Dotson et al. 2017) and sales (He et al. 2014).

The analysis combines data from TVision Insights and Google Trends. We select a sample of all 180 national ads aired during two NFL football playoff games on January 22, 2017, as they had large live audiences.[15]

Specifically, we ask the following questions: Can ad viewing measures predict brand search lift? How do empirical relationships differ between ad OTS and ad exposures? Answers may indicate how new viewing metrics relate to advertiser objectives.

5.1. Ad Data

Figure 15 illustrates four total viewing measures—OTS tuning, exposure tuning, presence, and attention—across the 180 ad slots, in chronological order with vertical lines between ad breaks. As before, ad OTS and ad exposures differ in the number of potentially exposed viewers who are absent during tuned ads. As expected, viewing typically decreases within breaks and increases between most breaks.

5.2. Search Volume Data

We downloaded minute-by-minute Google Trends reports for all brands advertised in the two games, using the main brand keyword for each ad. Google provides relative search volume indices, so every

Table 3. Descriptive Results of Pharmaceutical Category Causal Ad Effects on Drug Category Prevalence and Severity

Dependent variable	(1)	(2)	(3)
	Tuning Rx Effects		
Constant	−0.00454***	−0.00731***	−0.00389***
	(0.00001)	(0.00001)	(0.00001)
Severity	−0.00071***		−0.00072***
	(0.00000)		(0.00000)
Prevalence		−0.00037***	−0.00459***
		(0.00005)	(0.00003)
	Presence Rx Effects		
Constant	−0.00177***	−0.00333***	−0.00098***
	(0.00001)	(0.00001)	(0.00001)
Severity	−0.00046***		−0.00048***
	(0.00000)		(0.00000)
Prevalence		−0.00262***	−0.00564***
		(0.00005)	(0.00004)
	Attention Rx Effects		
Constant	−0.00075***	−0.00127***	−0.0001***
	(0.00000)	(0.00000)	(0.00000)
Severity	−0.00012		−0.0027***
	(0.00000)		(0.00001)
Prevalence		0.00018***	−0.0037***
		(0.00001)	(0.00003)

Notes. Columns (1) and (2) report results for severity and prevalence separately, whereas column (3) includes both. Standard errors are in parentheses.
***$p < 0.001$.

query also included the control keyword "Pizza Hut" to set a standard, comparable scale across Google Trends reports.

Figure 16 depicts the search data for 25 brand keywords from a variety of categories; full data are in Figure A8 in the online appendix. Baseline search levels and lifts varied across brands. Search sometimes spiked without an ad (e.g., Apple, Verizon), likely due to TV ads on other channels or in-game sponsor messages.

5.3. Model

Equation (3) relates brand search lift to viewing behaviors as follows:

$$LIFT_s = \eta Y_s^k + \phi Z_s + \varepsilon_s, \tag{3}$$

where s indexes the 180 ad slots, each of which represents a unique combination of brand b and minute t; $LIFT_s$ measures brand search in minutes t and $t+1$ minus brand search in minutes $t-1$ and $t-2$; and Y_s^k is one of the four total viewing measures, indexed by k, in slot s. The smallest correlation among total viewing measures is 0.83 (presence/attention), so including multiple viewing measures induces multicollinearity. The control vector, Z_s includes an intercept, a first-slot dummy, and, in some specifications, 22 quarter-hour dummies to control for time-varying drivers of advertiser targeting and brand search (e.g., tension within the game).

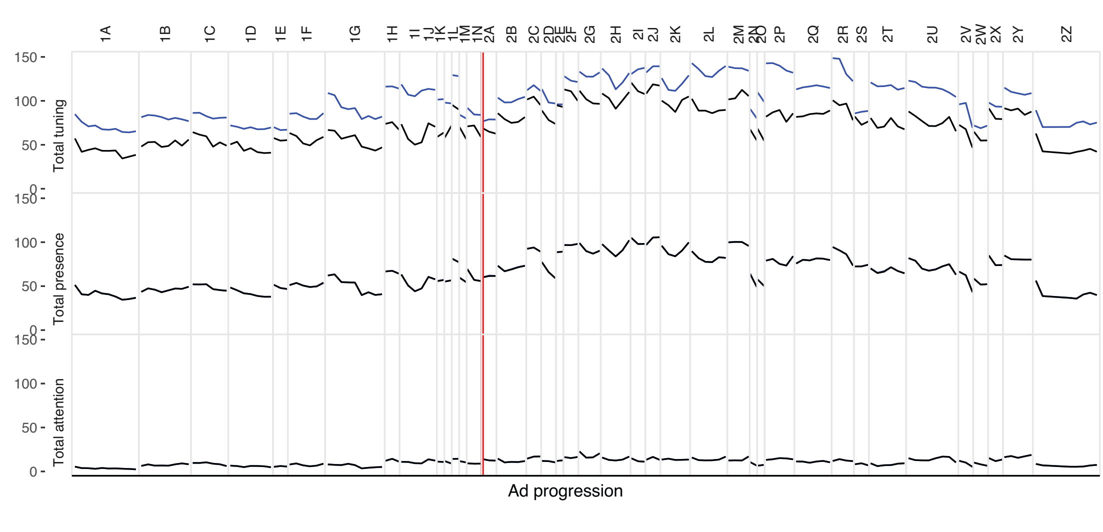
5.4. Results

Table 4 presents eight regressions, one for each of the four total viewing measures, with and without quarter-hour dummies. There are two key results:

1. OTS tuning does not reliably predict search lift. However, exposure tuning, presence, and attention all significantly predict search lift without quarter-hour fixed effects. Attention explains the most variation in search lift, followed by exposure tuning.

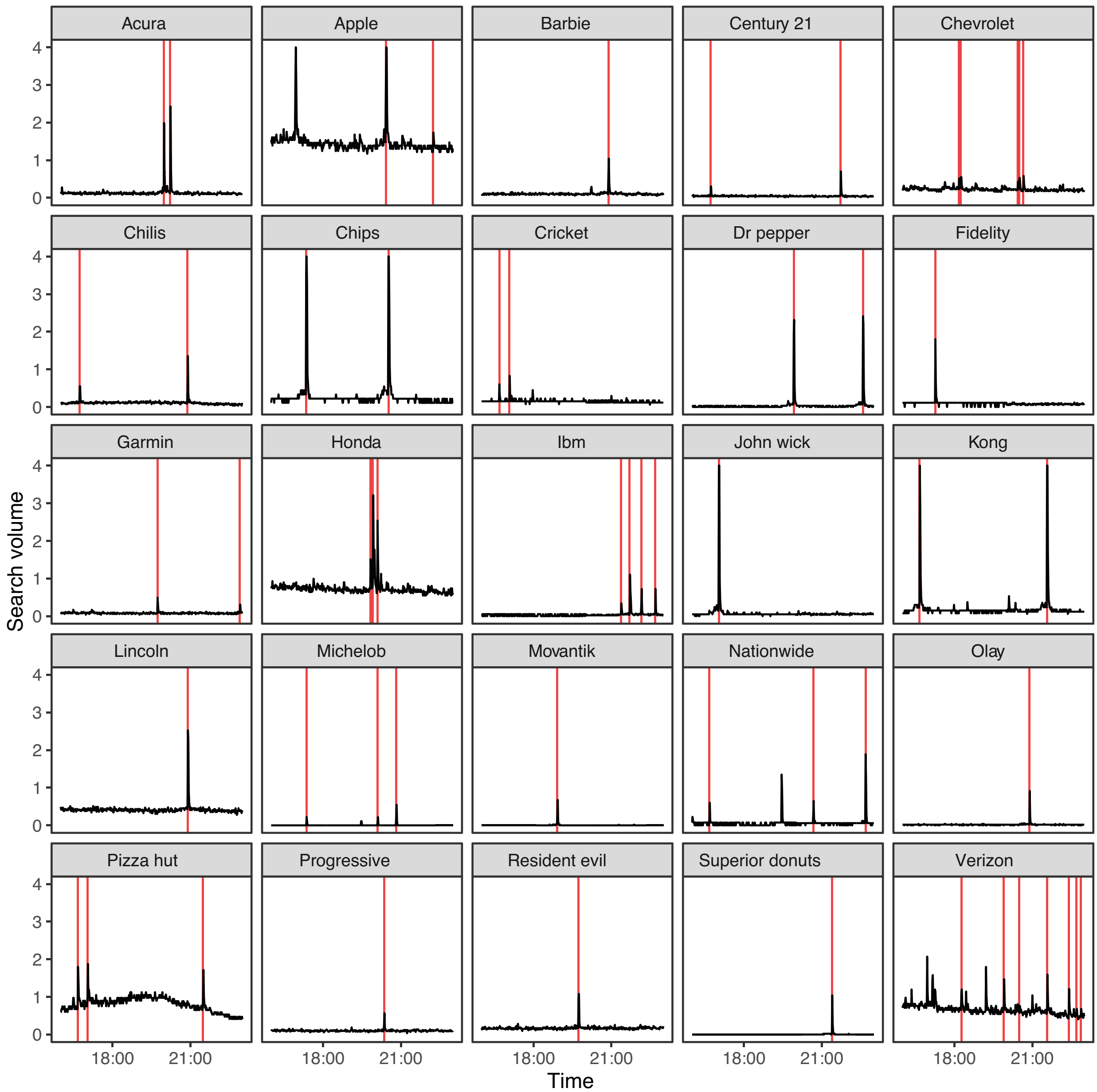
2. Attention significantly predicts search lift in the regressions with quarter-hour fixed effects.

Figure 15. Total Viewing Measures Across Two NFL Playoff Games in 2017



Notes. Panels show numbers of viewers tuned (in OTS and exposures), present, and attentive across consecutive ad slots in ad breaks with at least two national ads. Ad breaks are separated by gray vertical bars. The red line marks when the first game ends.

Figure 16. Examples of Brand Search Lift



Notes. Panels show Google search volume for 25 brands from a variety of categories, in units normalized to the average per-minute search for the keyword "Pizza Hut" in a reference hour, truncated at four. Red lines denote minutes brand ads began during the two NFL playoff games. The y-axis is truncated so as to highlight baseline variation.

5.5. Interpretation

The regressions suggest that advertisers can best explain search lift by measuring attention directly, or at least distinguishing between ad OTS and exposures. We conclude that the new viewing measures may help inform advertiser objectives.

We note two important caveats. First, this exercise is a small study of only two particular football games. A larger-scale study could yield more representative, better powered results, which may vary across brands, audiences, creatives, programs, or time. Second, we focused on brand search lift because of its convenient measurement across brands and time. Individual advertisers likely should consider proprietary response metrics (e.g., website visits, app usage, sales), as they would better indicate ad profits.[16]

6. Discussion and Conclusion

New TV ad viewing metrics offer the first passive, continuous, in situ measures of video ad exposures, showing that 30% of TV ads play to empty rooms. We

Table 4. Parameter Estimates for Dependent Variable *Search Lift*

Measure	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
OTS tuning	0.006				0.006			
	(0.005)				(0.01)			
Exposure tuning		0.015*				0.024		
		(0.006)				(0.013)		
Presence			0.013*				0.017	
			(0.006)				(0.014)	
Attention				0.064*				0.106*
				(0.029)				(0.052)
First ad fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quarter-hour fixed effects	No	No	No	No	Yes	Yes	Yes	Yes
R^2	0.035	0.082	0.07	0.116	0.158	0.176	0.165	0.180
N	180	180	180	180	180	180	180	180

Note. Standard errors are in parentheses.
$^*p < 0.05$.

constructed ad features and used a verifiably quasi-experimental identification strategy to estimate how ad content influences ad viewing. We find that recreational product categories tend to preserve tuning and presence, whereas prescription drug ads tend to reduce tuning and presence, especially for prevalent and severe conditions. Attention falls during longer ads and early in commercial breaks. A supplementary, spot-level analysis finds that attention predicts brand search lift better than traditional metrics. Taken together, our analyses suggest that these new metrics may have important implications for advertisers, ad-selling platforms, and consumers.

6.1. Implications for Advertisers

Brand advertisers typically do not have granular outcome data to discipline their campaign choices, raising the question of whether attention data could serve as an intermediate success metric. Performance advertisers, on the other hand, can directly test the empirical value of the new viewing metrics. The search lift analysis suggests that measuring attention and distinguishing OTS from exposures may both help predict ad response and optimize media buying.

Our analysis also draws attention to the fact that viewing metrics may have to be filtered carefully between targeted and nontargeted ad viewers to fully understand return on advertising investment. For example, the pharmaceutical category results summarize how the mass audience reacts to drug ads, but they do not imply that drug ads do not work, that is, a cancer patient likely reacts quite differently to a cancer drug ad than the median viewer.

6.2. Implications for Platforms and Viewers

Ads that reduce viewing decrease subsequent ads' audience. In general, video ad attention depends on program environment and earlier ads (Webb 1979, Burke and Srull 1988, Joo et al. 2020, Liu et al. 2021, Rajaram et al. 2021). Closely related evidence exists in mobile advertising (Rafieian and Yoganarasimhan 2020) and search advertising (Gomes et al. 2009). Importantly, highly involving ads also may help retain audience in subsequent ad slots (Fossen et al. 2020).

Media platforms—for example, Hulu, NBC, *New York Times*, YouTube—already restrict ad content to help preserve viewership and brand safety. Our analysis suggests there is further scope to improve the design of TV ad markets by incorporating attention-related metrics. This idea has been explored in a related literature that analyzes how ad markets should select, order, and price ads to internalize short-run negative externalities across ads (Wilbur et al. 2013, Kar et al. 2015, Theocharous et al. 2015, Deng and Mela 2018, Rafieian 2019, Rajaram et al. 2021).

Attention preservation also has long-run implications. In particular, it may affect the number and types of viewers attracted to a platform, and the habitual ad responses they develop. Incorporating attention metrics in ad market design could be an efficiency improvement benefiting all sides of this three-sided market.

6.3. Limitations and Extensions

The current paper has several important limitations, as does all research. We relied exclusively on quasi-random variation to estimate causal effects, but that may have over-controlled for endogeneity. We estimated local average treatment effects rather than average treatment effects or heterogeneous treatment effects. We presume that consumers often divert attention during ads because of second-screening behaviors, but we are not yet able to measure such behaviors directly. We have not investigated the implications of coviewing, ad frequency, or other related topics.

Many opportunities remain unexplored. TV ad field experiments are too rare. Randomization could help identify effects of ad insertion, targeting, and content on viewing metrics and sales. Further, no one has directly quantified how presence or attention influence conversions, or vice versa. Another interesting avenue is to quantify advertising return on investment, focusing on whether award-winning ads, or ads from prestigious or expensive ad agencies, perform better on these or similar measures.[17] Whereas our data are not well equipped to answer this question, we think this is an interesting avenue for future research.

In summary, new viewing metrics differ meaningfully from traditional metrics, respond to ad content, and help predict ad response. There has been limited competition among TV ad measurement providers in the past, but improved viewer presence and attention measures could improve video ad transactions and overall market efficiency.

Acknowledgments
The authors thank TVision Insights, especially Tristan Webster, Yan Liu, and Dan Schiffman, for enabling this research and providing valuable feedback. The authors also thank Thales Teixeira at Decoupling.co for connecting them with TVision, Snehanshu Tiwari, and Haoyue Yang for superlative research assistance, and seminar participants at Bocconi University, Boston University, Cornell University, University of Delaware, Georgetown University, University of Indiana, University of Minnesota, University of California, San Diego, University of Texas, Dallas, Virginia Tech, University of Western Ontario, and University of Wisconsin for helpful suggestions. An earlier version dated July 2019 was called "Watching People Watch TV." Mistakes, omissions, and opinions are the authors' alone.

Endnotes

[1] This is only one possible measure of attention. A viewer may actively process ad audio while looking away from the screen, or stare at the screen yet be entirely absorbed in other thoughts, or focus on a program but blinking or head movements may lead to average attention well below 100%.

[2] We select 50 to minimize sampling variance from infrequent guests or mistaken person classifications.

[3] We checked the TVision ad insertion data against the official advertising schedule for Super Bowl 51 and against Kantar Media, a reliable commercial source of ad insertion data. The TVision ad data contained all 63 national ads in the correct order. The average insertion time difference was 4.9 seconds, consistent with standard asynchronies in local broadcast affiliate streams. We also found a very high correspondence between TVision ad insertion data and Kantar data in other programs.

[4] In the sample, regular viewers—defined as those with at least 50 active viewing days—view 22.5 sample ads per person per day and attend to 1.1 out of 8.9 exposed ad minutes.

[5] We checked but did not find extensive collinearity within iSpot features or within machine learning-coded features.

[6] Blanchard et al. (2019) provide a detailed overview of the video coding platform and find that image tags help to predict new product adoption. Kubany et al. (2020) found Google Cloud Vision performs well compared with competing image recognition services.

[7] Equation (1) implicitly assumes that any viewer can change device-level tuning. See Yang et al. (2010) and references therein for an alternate approach of group-level decision making. Figure A6 in the online appendix shows that main findings are unchanged when the sample is restricted to single-viewer sessions.

[8] We decided against using a discrete choice model because (i) response behaviors are continuous and (ii) choice sets are unobserved but vary across viewers and viewing sessions, for example, during live or time-shifted viewing.

[9] In our sample, 3.5% of ad exposures occur during viewer-break combinations with a single ad exposure, when viewer-break fixed effects are inseparable from ad effects. Ad effects are identical whether we drop or retain single-slot viewer-break combinations.

[10] Figure 5 excludes promotions for network programs, which often appear before or after commercial breaks.

[11] Those include studies of TV ad avoidance (Wilbur et al. 2013); brand website visits and sales (Liaukonyte et al. 2015, He and Klein 2018, Meder et al. 2019); social media chatter (Fossen and Schweidel 2019); brand search and price search (Du et al. 2019); subsequent TV ads' audience and resulting digital chatter (Fossen et al. 2020); and brand awareness, consideration, and purchase (Tsai and Honka 2021).

[12] We considered using inverse probability weighting (IPW), as in single-treatment settings (Gordon et al. 2019, Rafieian and Yoganarasimhan 2020). However, there is no consensus on how best to implement IPW in settings with either multivariate treatments or multiple treatments (Lopez and Gutman 2017); our setting features both.

[13] Limited absolute changes could still be appreciable on a relative basis given that average attention is near 13%.

[14] Alternatively, drug ads for more prevalent conditions may be more relevant to a wider set of viewers, but we question whether those viewers would prefer to receive such messages during television program consumption.

[15] Twenty-one percent of TVision panelists were exposed to the ads. For comparison, Nielsen estimated audience ratings of 24.4 and 25.0 for the two games (Sports Media Watch 2017). The differences likely reflect sample selection and measurement differences between TVision and Nielsen, as well as sampling error. We do not analyze Super Bowl data because many viewers gather specifically to watch Super Bowl ads, a nonmodal behavior.

[16] In particular, if a brand advertises on its own keyword, it may have to pay Google a significant toll (Simonov and Hill 2021) and some resulting searches could be accelerated in time rather than incrementally (Lambrecht et al. 2021).

[17] Toward this goal, we estimated a version of the model including an indicator of TV ad awards (Clio, Effie, or Emmy). The award dummy was directionally positive but not statistically distinguishable from zero. This analysis was likely underpowered as there were fewer than 20 award-winning creatives in our sample, totaling around 5,000 impressions.

References

Anderson SP, Ciliberto F, Liaukonyte J (2013) Information content of advertising: Empirical evidence from the OTC analgesic industry. *Internat. J. Indust. Organ.* 31(5):355–367.

Anderson SP, Ciliberto F, Liaukonyte J, Renault R (2016) Push-me pull-you: Comparative advertising in the OTC analgesics industry. *RAND J. Econom.* 47(4):1029–1056.

Blanchard S, Noseworthy TJ, Pancer E, Poole M (2019) Visual ambiguity in product design predicts crowdfunding success. *ACR North American Advances*.

Bronnenberg BJ, Dube J-P, Mela CF (2010) Do digital video recorders influence sales? *J. Marketing Res.* 47:998–1010.

Burke RR, Srull TK (1988) Competitive interference and consumer memory for advertising. *J. Consumer Res.* 15(1):55–68.

Chandrasekaran D, Srinivasan R, Sihi D (2018) Effects of offline ad content on online brand search: Insights from Super Bowl advertising. *J. Acad. Marketing Sci.* 46:403–430.

Chintagunta PK, Nair HS (2011) Structural workshop paper-discrete-choice models of consumer demand in marketing. *Marketing Sci.* 30(6):977–996.

Danaher PJ (1995) What happens to television ratings during commercial breaks? *J. Advertising Res.* 35(1):37–47.

Deng Y, Mela CF (2018) TV viewing and advertising targeting. *J. Marketing Res.* 55(1):99–118.

Dotson JP, Fan RR, McDonnell Feit E, Oldham JD, Yeh Y-H (2017) Brand attitudes and search engine queries. *J. Interactive Marketing* 37:105–116.

Du RY, Xu L, Wilbur KC (2019) Immediate responses of online brand search and price search to TV ads. *J. Marketing* 83(4):81–100.

eMarketer (2019) US TV ad spending, by type, 2017–2021. Accessed October 25, 2021, https://web.archive.org/web/20220113185257/https://www.emarketer.com/chart/231776/%20us-tv-ad-spending-by-type-2017-2021-billions-change.

Ephron E (2006) Who watches the watchers? Accessed December 15, 2020, http://web.archive.org/web/20201218140310/https://insightandresearch.wordpress.com/2006/11/13/who-watches-the-watchers/.

Food and Drug Administration (2020) Prescription drug advertising. Accessed December 15, 2020, http://web.archive.org/web/20210105200414/https://www.fda.gov/drugs/prescription-drug-advertising/drug-advertising-glossary-terms.

Fossen BL, Schweidel DA (2019) Social TV, advertising, and sales: Are social shows good for advertisers? *Marketing Sci.* 38(2):274–295.

Fossen BL, Mallapragada G, De A (2020) Impact of political television advertisements on viewers' response to subsequent advertisements. *Marketing Sci.* 40(2):305–324.

Friedman W (2017) Shorter-duration TV commercials on the rise. *Television News Daily* (October 4), https://web.archive.org/web/20210127013859/https://www.mediapost.com/publications/article/308248/shorter-duration-tv-commercials-on-the-rise.html.

Goettler RL, Shachar R (2001) Spatial competition in the network television industry. *RAND J. Econom.* (Springer, Berlin Heidelberg), 32(4):624–656.

Gomes R, Immorlica N, Markakis E (2009) Externalities in keyword auctions: An empirical and theoretical assessment. Leonardi S, ed. *Proc. Internat. Workshop Internet Network Econom.* (Springer, Berlin Heidelberg), 172–183. https://link.springer.com/chapter/10.1007/978-3-642-10841-9_17.

Gordon BR, Zettelmeyer F, Bhargava N, Chapsky D (2019) A comparison of approaches to advertising measurement: Evidence from big field experiments at Facebook. *Marketing Sci.* 38(2):193–225.

Gordon BR, Jerath K, Katona Z, Narayanan S, Shin J, Wilbur KC (2021) Inefficiencies in digital advertising markets. *J. Marketing* 85:7–25.

Guimaraes P, Portugal P (2010) A simple feasible procedure to fit models with high-dimensional fixed effects. *Stata J.* 10(4):628–649.

He C, Klein T (2018) Advertising as a reminder: Evidence from the Dutch state lottery. Working paper, ShanghaiTech University–School of Entrepreneurship and Management.

He Y, Du RY, Damangir S (2014) Decomposing the impact of advertising: Augmenting sales with online search data. *J. Marketing Res.* 51:300–319.

Iizuka T, Jin GZ (2005) The effect of prescription drug advertising on doctor visits. *J. Econom. Management Strategy* 14(3):701–727.

Iizuka T, Jin GZ (2007) Direct to consumer advertising and prescription choice. *J. Indust. Econom.* 55(4):771.

Imbens GW, Angrist JD (1994) Identification and estimation of local average treatment effects. *Econometrica* 62(2):467–475.

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

Joo M, Liu W, Wilbur KC (2020) Divergent temporal courses for liking vs. wanting in response to persuasion. *Emotion* 20(2):261–270.

Kar W, Swaminathan V, Albuquerque P (2015) Selection and ordering of linear online video ads. *Proc. 9th ACM Conf. Recommender Systems, Vienna, Austria* (Association for Computing Machinery, New York), 203–210.

Knauer L (2019) Revised MRC viewability guidelines: Everything you need to know. Accessed March 26, 2021, https://web.archive.org/web/20210902195246/https://www.brightcove.com/en/resources/blog/revised-mrc-viewability-guidelines-everything-you-need-know/.

Kubany A, Ishay SB, Ohayon RS, Shmilovici A, Rokach L, Doitshman T (2020) Comparison of state-of-the-art deep learning APIs for image multi-label classification using semantic metrics. *Expert Systems Appl.* 161:113656.

Lambrecht A, Tucker CE, Zhang X (2021) Does TV advertising increase online sales: The role of inter-temporal substitution. Working paper, London Business School.

Lee D, Hosanagar K, Nair HS (2018) Advertising content and consumer engagement on social media: Evidence from Facebook. *Management Sci.* 64(11):5105–5131.

Lewis RA, Rao JM (2015) The unfavorable economics of measuring the returns to advertising. *Quart. J. Econom.* 130(4):1941–1973.

Liaukonyte J, Žaldokas A (2021) Background noise? TV advertising affects real-time investor behavior. *Management Sci.*, ePub ahead of print August 10, https://doi.org/10.1287/mnsc.2021.4003.

Liaukonyte J, Teixeira T, Wilbur KC (2015) Television advertising and online shopping. *Marketing Sci.* 34(3):311–330.

Liu X, Shum M, Uetake K (2021) Passive and active attention to baseball telecasts: Implications for content (re-)design. Working paper, New York University (NYU)–Leonard N. Stern School of Business.

Liu X, Shi SW, Teixeira T, Wedel M (2018) Video content marketing: The making of clips. *J. Marketing* 82(July):86–101.

Lopez MJ, Gutman R (2017) Estimation of causal effects with multiple treatments: A review and new ideas. *Statist. Sci.* 32(3):432–454.

Mandese J (2004) Equitable's no longer questionable: Data reveals nets position some advertisers better than others. *Media Daily News* (September 17), https://web.archive.org/web/20140829183611/https://www.mediapost.com/publications/article/18344/equitables-no-longer-questionable-data-reveals-n.html.

Meder CJ, Kemper J, Brettel M (2019) How creative cues determine online response to TV commercials: An empirical investigation for digital-native brands. vom Brocke J, Gregor S, Muller O, eds. *Proc. 27th European Conf. on Information Systems for a Sharing Society, (ECIS) 2019, Stockholm and Uppsala, Sweden, June 8–14*.

Rafieian O (2019) Optimizing user engagement through adaptive ad sequencing. Working paper, SC Johnson School of Management, Cornell University, New York.

Rafieian O, Yoganarasimhan H (2020) How does variety of previous ads influence consumer's ad response? Working paper, SC Johnson School of Management, Cornell University, New York.

Rajaram P, Manchanda P, Schwartz EM (2021) Finding the sweet spot: Ad targeting on streaming media. Working paper, Ivey Business School, Western University, London.

Resnik A, Stern BL (1977) An analysis of information content in television advertising. *J. Marketing* 41(1):50–53.

Schweidel DA, Kent RJ (2010) Predictors of the gap between program and commercial audiences: An investigation using live tuning data. *J. Marketing* 74(3):18–33.

Schwenzow J, Hartmann J, Schikowsky A, Heitmann M (2021) Understanding videos at scale: How to extract insights for business research. *J. Bus. Res.* 123:367–379.

Shachar R, Emerson JW (2000) Cast demographics, unobserved segments, and heterogeneous switching costs in a television viewing choice model. *J. Marketing Res.* 37(2):173–186.

Shapiro BT (2018) Positive spillovers and free riding in advertising of prescription pharmaceuticals: The case of antidepressants. *J. Political Econom.* 126(1):381–437.

Shapiro B (2022) Promoting wellness or waste? Evidence from antidepressant advertising. *Amer. Econom. J. Microeconom.* Forthcoming.

Shapiro B, Hitsch GJ, Tuchman A (2021) TV advertising effectiveness and profitability: Generalizable results from 288 brands. *Econometrica* 89(4):1855–1879.

Siddarth S, Chattopadhyay A (1998) To zap or not to zap: A study of the determinants of channel switching during commercials. *Marketing Sci.* 17(2):124–138.

Simonov A, Hill S (2021) Competitive advertising on brand search: Traffic stealing and click quality. *Marketing Sci.* 40(5):923–945.

Sinkinson M, Starc A (2018) Ask your doctor? Direct-to-consumer advertising of pharmaceuticals. *Rev. Econom. Stud.* 86(2):836–881.

Sports Media Watch (2017) 2016 NFL TV ratings. Accessed August 10, 2021, https://web.archive.org/web/20210103192028/https://www.sportsmediawatch.com/2016-nfl-tv-ratings-viewership/.

Steinberg B (2021) Nielsen's national ratings accreditation suspended by industry group. *Variety* (September 1), https://web.archive.org/web/20210902184602/https://variety.com/2021/tv/news/nielsen-national-ratings-suspension-media-rating-council-1235053949/.

Swaminathan S, Kent R (2013) Second-by-second analysis of advertising exposure in TV pods: The dynamics of position, length, and timing. *J. Advertising Res.* 53(1):91–100.

Swant M (2019) Results of IAB Tech Laboratory's blockchain pilot shows how many digital ads are seen by humans. Accessed March 26, 2021, https://web.archive.org/web/20210117165610/https://www.adweek.com/performance-marketing/results-of-iab-tech-labs-blockchain-pilot-shows-how-many-digital-ads-are-seen-by-humans/.

Teixeira TS, Wedel M, Pieters R (2010) Moment-to-moment optimal branding in TV commercials: Preventing avoidance by pulsing. *Marketing Sci.* 29(5):783–804.

Teixeira T, Wedel M, Pieters R (2012) Emotion-induced engagement in Internet video advertisements. *J. Marketing Res.* 49 (2):144–159.

Theocharous G, Thomas PS, Ghavamzadeh M (2015) Personalized ad recommendation systems for life-time value optimization with guarantees. Yang Q, ed. *Proc. Twenty-Fourth Internat. Joint Conf. Artificial Intelligence* (AAAI Press, Palo Alto, CA), https://www.ijcai.org/Proceedings/2015/.

Tsai Y-L, Honka E (2021) Informational and non-informational advertising content. *Marketing Sci.* 40(6):1030–1058.

Tuchman AE, Nair HS, Gardete PM (2018) Television ad-skipping, consumption complementarities and the consumer demand for advertising. *Quant. Marketing Econom.* 16(2):111–174.

Tucker CE (2015) The reach and persuasiveness of viral video ads. *Marketing Sci.* 34(2):281–296.

Voorveld HA, Viswanathan V (2015) An observational study on how situational factors influence media multitasking with TV: The role of genres, dayparts, and social viewing. *Media Psych.* 18(4):499–526.

Webb PH (1979) Consumer initial processing in a difficult media environment. *J. Consumer Res.* 6(3):225–236.

Wilbur KC (2008) How the digital video recorder (DVR) changes traditional television advertising. *J. Advertising* 37(1):143–149.

Wilbur KC (2016) Advertising content and television advertising avoidance. *J. Media Econom.* 29(2):51–72.

Wilbur KC, Xu L, Kempe D (2013) Correcting audience externalities in television advertising. *Marketing Sci.* 32(6):892–912.

Yang S, Zhao Y, Erdem T, Zhao Y (2010) Modeling the intrahousehold behavioral interaction. *J. Marketing Res.* 47(3):470–484.

Zhang J, Wedel M, Pieters R (2009) Sales effects of attention to feature advertisements: A Bayesian mediation analysis. *J. Marketing Res.* 46(5):669–681.

Zigmond D, Stipp H (2010) Assessing a new advertising effect: Measurement of the impact of television commercials on Internet search queries. *J. Advertising Res.* 50(2):162–168.

Zufryden FS, Pedrick JH, Sankaralingam A (1993) Zapping and its impact on brand purchase behavior. *J. Advertising Res.* 33(1):58–67.