Permission Advertising: Understanding Pre-Roll Ads Leveraging Artificial Empathy

Xinying Hao* Garrett Sonnier† Frenkel ter Hofstede†

Abstract

Ad avoidance behavior is an increasingly important problem for social platforms that rely on permission advertising revenues. A common tactic employed by platforms is to impose a period of forced ad exposure. For example, YouTube typically requires viewers to watch the first five seconds of an ad, after which the viewer can choose to skip the ad and proceed to the desired content. In this paper, we develop a model that quantifies the effects of forced ad exposure on consumers' emotions and ad-skipping behavior when watching online video advertisements. We use *artificial empathy* (i.e., facial recognition technology) to measure emotions in a way that is completely unobtrusive, thus avoiding mere measurement effects. Leveraging state-of-the-art computer vision techniques, we also extract frame-level features from video advertisements, such as canny edges and the foreground. Our Bayesian dynamic generalized linear model captures the temporal trajectory of consumer emotions under forced and unforced ad exposure conditions as well as the dynamics of the consumer's ad skipping behavior. Our results indicate that forced ad exposure largely ignites contempt and disgust and suppresses feelings of surprise. Surprise and anger cause a decrease in skipping probability while contempt, disgust, and sadness increase the risk of losing the audience's attention. Moreover, we find a high carryover effect in the skipping propensity, which highlights the importance of capturing viewers' attention in the opening seconds. Our study provides practical insights for ad designers, advertisers, and online video platforms.

Keywords: Online Skippable Ads, Forced Ad Exposure, Ad Avoidance, Artificial Empathy, Dynamic Generalized Model

*Assistant Professor of Marketing at Robert H. Smith School of Business, University of Maryland, College Park.
†Associate Professor of Marketing at the McCombs School of Business, University of Texas at Austin.
This is a preliminary draft. Please do not circulate.

1 Introduction

To enhance the user experience and minimize serving advertisements to uninterested viewers, digital platforms have progressively adopted "permission advertising". Permission advertising refers to endowing consumers with some agency in the decision of whether or not to engage with the advertisement. Consumers have long had at least some degree of agency, be it changing the radio station or flipping the television channel. However, in these settings consumers could not choose to skip the advertisement to reach the next song or program on a given channel. Online video platforms provide exactly this agency via the skippable ad format. This approach empowers online viewers to bypass unwanted ads with a simple click, significantly improving their interaction with online content. This has proven to be a profitable strategy (Dukes et al., 2019; Dukes and Liu, 2023). For instance, Google's acquisition of YouTube in 2006 for $1.65 billion initially yielded little revenue, but with the introduction of skippable ads in 2010, the platform's revenue accelerated (WSJ, 2015). Therefore, ad avoidance behavior has become an increasingly crucial issue for social platforms relying on permission advertising revenues.

The internet has not only revolutionized the way ads can be skipped, but also transformed how the resulting ad-skipping behaviors are valued. Three key points highlight this shift. First, in the era of the "Skip Ad" button, video streaming platforms still possess the authority to mandate a period of ad exposure prior to allowing the skip decision. For example, YouTube typically requires viewers to watch the first five seconds of an ad, after which the viewer can choose to skip the ad and proceed to the desired content. Second, online ad-skipping, in contrast to channel switching, can yield more immediate favorable outcomes for viewers, underlining the significant role of emotions in this process. By simply clicking the "Skip Ad" button, online viewers can swiftly bypass ads and proceed to watch videos of their choosing without the frustration of repeatedly using the "back" button on a remote control during a commercial break (Globe, 2013). Neuroscientific research has discovered that in decisions involving immediate outcomes, such as whether to skip an ad or not, areas of the brain closely linked to emotions become activated (McClure et al., 2004). Third, how long online viewers are actually watching ads is a central concern for both advertisers and online networks. From the platforms' perspective, if skippable ads fail to retain viewers, advertisers may only be required to pay a fraction or none of the advertising fees. For instance, according to

YouTube's TrueView model, advertisers are charged only when viewers watch 30 seconds of the ad or its entirety if the ad is shorter than 30 seconds. This approach differs from traditional cost-per-click or cost-per-thousand impressions pricing, as advertisers are now billed based on viewers' watching durations. In general, permitting ad-avoidance behavior serves as a method for pricing customers' time and attention. Given the changes in online advertising and ad-avoidance tendencies among online viewers, there is a growing need to understand the factors why and when online viewers decide to skip ads.

Our research centers on the effects of forced ad exposure on skipping behavior in online ads. We conjecture that the initial moments of forced ad exposure may exert both positive and negative influences on viewer behavior. On the positive side, the content featured in the first few seconds of the ad, especially during the forced viewing period, significantly shapes the viewer's subsequent decision to skip the ad. As indicated by Aravindakshan and Naik (2011), the carry-over effect of advertising on consumer awareness underscores the critical impact of these initial ad moments on future viewer actions. The work of Hart (2015) highlights that the presence of a recognizable celebrity within the first five seconds of a skippable ad often correlates with increased viewership and improved brand lift. Additionally, according to Hart (2015), certain musical styles—such as calming, relaxing, or action-oriented compositions—proves more effective in the first five seconds compared to others. This indicates that forced ad exposure offers advertisers an opportunity to establish the right tone and captivate their audience in the initial moments. Consequently this alters the distribution of viewers' subsequent engagement with the ad. However, forced ad exposure may also yield adverse effects. The intrusive nature of such exposure can trigger negative emotions and diminish an individual's sense of freedom and autonomy, undermining the effectiveness of the advertisement (Edwards et al., 2002). For instance, there have been growing complaints on platforms like YouTube that cite potential waste of advertisers' funds when ads are not visible to viewers who minimize their web browsers during forced exposure. This effectively leads to ad skipping, but in this case the skipping behavior is opaque (FinancialTimes, 2015). Thus, in understanding ad-skipping behavior, the significance of the opening moments of forced ad exposure cannot be overlooked.

In this paper, we study the effects of forced ad exposure in the context of a skipable online ad. In particular, we focus on the role of emotions. A challenge in measuring emotions is obtrusiveness which may invite mere measurement effects. We utilize unobstrusive facial tracking techniques (i.e. *A*rtificial Empathy) which allow us to navigate around mere measurement effects. We utilize a model-based approach to quantify the effects of forced ad exposure on consumer emotions and subsequent ad-skipping behavior when watching online video advertisements. More specifically, we employ a Bayesian generalized dynamic linear model to address the following research questions: (1) How does forced ad exposure affect emotions and emotional trajectories? (2) Conditional on forced ad exposure, what is the effect of emotions on ad-skipping behavior? (3) For an online video platform, what is the most effective forced watching time in terms of maximizing viewership? (4) For advertisers/ad designers, how best to evoke emotional appeal in an ad? Our results indicate that forced ad exposure predominantly triggers contempt and disgust while simultaneously diminishing the feeling of surprise. Surprise and happiness significantly reduce the probability of skipping an ad. Contempt, disgust, and sadness increase the risk of losing the audience's attention. The substantial carryover effect on skipping propensity underscores the importance of captivating viewers within the initial seconds. Our study's findings have implications for ad designers, advertisers, and online video platforms.

The remainder of the paper is organized as follows. Section 2 introduces the role of emotions. Section 3 outlines artificial empathy and the experimental design. Section 4 presents the Bayesian Dynamic Linear Model. Sections 5 and 6 discuss our data and estimation results. Finally, Section 7 elucidates the implications of our findings and concludes the study with limitation and future research.

2 Literature Review

2.1 Ad Avoidance

Our paper contributes to the research exploring ad-skipping behavior in the presence of forced ad exposure. Prior research on ad-skipping behavior has focused on identifying cognitive factors that affect ad-skipping decisions. For example, Schweidel and Kent (2010) and Kent and Schweidel (2011) utilize set-top box tuning data to analyze variations in program audiences across programs and advertisements within commercial breaks. They find that ad-avoidance behavior is moderated by the specific program and advertisements' position within a break but are unable to discern individual-level skipping behaviors. Siddarth and Chattopadhyay (1998) leverage household-level commercial viewing panel data to outline crucial factors driving channel switching during commercials, including purchase history, ad familiarity, commercial length, content, and timing within a commercial break. By employing eye-tracking technology, Teixeira et al. (2010) develop a conceptual model to explain how branding activities affect consumer moment-to-moment avoidance decisions during television advertising.

While extant research explores the impact of cognitive factors on live and recorded television advertising, it notably lacks insights into how the affective system (i.e., emotions) influences ad-skipping decisions. In addition, there is scant empirical research in context of online video advertising with interventions from online video platforms. The affective system is likely to play an important role in short-term inter-temporal decisions, such as ad-skipping. Chang and Pham (2013) argue that there is a greater reliance on affective feelings in judgments and decisions whose outcomes or targets are closer to the present. Furthermore, what effectively applies to traditional commercials may not seamlessly translate to online ads, particularly concerning viewers being compelled to watch an online ad. To elucidate the impact of the affective system on moment-to-moment skipping decisions, Texeira et al. (2012) use facial recognition technology to show how advertisers can use joy and surprise to engage consumers in watching online video ads. However, this study does not consider forced ad exposure. There exists a gap in understanding how viewers' skipping propensity and emotions evolve under forced ad exposure. We fill this gap by studying the dynamics of skipping propensity and emotional evolution in the context of forced ad exposure within a digital advertising setting. Moreover, our approach offers a novel way to identify potential nontransparent skipping behavior during the forced watching period, such as minimizing the web browser or opening a new tab.

2.2 Emotions and Facial Expressions

Understanding emotions is critical to studying ad-skipping behavior. We rely on two prominent models to assess emotions, the dimensional and categorical models. The dimensional model posits that emotions can be classified on a dimensional basis, predominantly based on arousal and valence. The circumplex model, as proposed by Russell (1980), suggests that emotions are distributed in a two-dimensional circular space: pleasure versus displeasure horizontally and arousal versus sleep vertically. This model's validity is supported by studies such as Abelson and Sermat (1962), demonstrating the similarity between expressions based on their geometric representation. In contrast, the categorical model relies on the fact that all humans are thought to have an innate set of six basic emotions– happiness, sadness, surprise, fear, anger, and disgust. Therefore emotions are discrete and fundamentally different constructs. Ekman et al. (1987) use a cross-cultural study to conclude that the six basic emotions are universal and cross-culturally recognizable. This theory proposes that facial expressions are innate and universal across cultures, forming part of a small set of "affect programs" associated with each universal emotion. Recent research and facial coding systems are aligned with these six universal emotions.

In the domain of affect, both models offer classifications of emotion. However, the existing literature lacks a consensus on which model better represents human emotions in decision-making. The categorical model offers more intuitive conclusions for designers and advertisers, given the extensive literature on transforming perceived emotions into advertising technologies. For instance, results embedded in a dimensional model are more complex and harder to explain. In contrast, existing research posits how to use advertising (e.g. humor scene) to induce categorical emotions (e.g., surprise) (Elpers et al., 2004). Our research integrates both models to investigate the impact of forced ad exposure on emotions. We establish more in-depth substantial implications using the six universal emotions model. Additionally, we include contempt—a nuanced blend of disgust and anger—as a seventh emotion due to its potential impact on ad-skipping behavior. Preliminary evidence supports the notion that contempt and its expression are universally recognized (Ekman and Friesen, 1969).

A novel aspect of our research is the method we employ to measure emotions. We measure emotions by tracking viewers' in a completely unobtrusive way. Despite the broad interest in emotions, much of the existing research in this area is limited. In order to measure emotions, three general classes of measurement methods have been proposed. First, because of its ease of use and versatility, self-report measures enjoy wide popularity and are the most common type of emotion measure (Laurans, 2011). Relying solely on a participant's memory and ability to recall a complex sequence of events, however, provides only insights into the ad as a whole and risks introducing additional biases. For instance, Alden et al. (2000) use surveys of retrospective self-reported surveys to obtain perceived humor measures. While such measurements are relevant for many studies that look at emotions in a static sense, they do not capture the dynamic nature of affect processes which may be especially important for ad-skipping behavior. Elpers et al. (2004) developed a slider tool and asked participants to move a mouse on a continuous scale to measure humorous emotions but the need for a quick response time largely reduced self-reporting accuracy. With the help of a slider tool, moment-to-moment measurements can be captured, but noise still exists and introduces measurement errors. Second, while the aforementioned research measures emotions obtrusively, other studies try to avoid interruptions by unobtrusively videotaping the respondents and relying on human coders (Ward and Broniarczyk, 2011). However, human coders are often potentially unreliable and unable to recognize so-called micro-expressions (i.e. a brief, involuntary facial expression shown on the face of humans according to emotions experienced). Third, in order to better capture these micro-expressions, researchers have recently proposed the use of face-tracking techniques to measure emotions. Teixeira et al. (2014) conduct a lab experiment and a field study, where participants were asked to turn on a camera so that their emotions could be measured. However, awareness of the measurement may change an individual's emotions and behaviors due to the mere measurement effect (Morwitz and Fitzsimons, 2004). Thus, we aim to measure emotions in a completely unobtrusive way.

Facial expressions, while not a direct one-to-one representation of emotions, provide reasonably accurate moment-to-moment measures of emotions. Facial expression analysis using tools like FaceReader shows a high accuracy rate (over 90%) compared to other existing measurements. Facial expressions also serve as a critical means of non-verbal communication, often revealing subconscious emotions that may not be explicitly expressed (Pease and Pease, 2008). As a primary form of communication learned from infancy, facial expressions provide essential clues for understanding human emotions. Therefore, our research endeavors to address these challenges by employing a technique that merges computer science's facial recognition capabilities with psychology. This technique offers a means to measure emotions unobtrusively, avoiding the biases introduced by traditional measurement methods. In the following section, we will detail our lab experiment design, leveraging these facial recognition techniques to measure emotions in an unobtrusive and accurate manner.

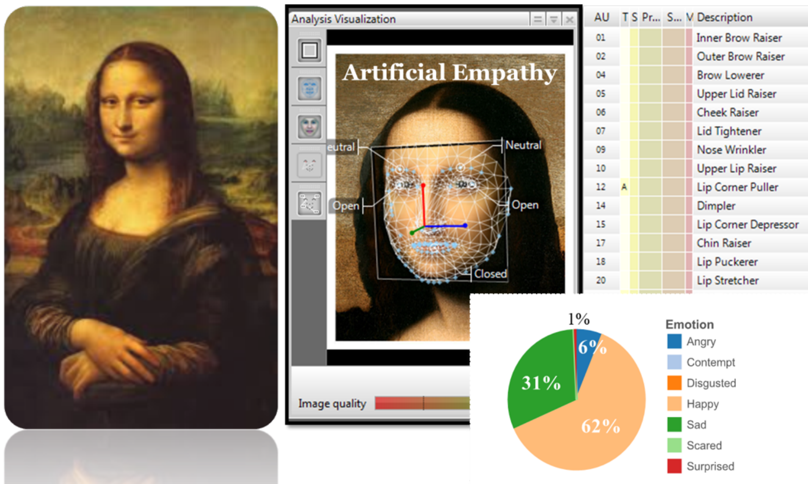


Figure 1: Examples of Facial Recognition

3 Facial Tracking Experiment

3.1 Artificial Empathy

Artificial Empathy (AE) is defined as the ability of nonhuman models to predict a person's internal state (e.g. affective) given the signals he or she emits (e.g., facial expression, voice, gesture) or to predict a person's reaction when he or she is exposed to a given set of stimuli (Xiao et al., 2013). In our study, we use a type of AE that relies on a technology developed in computer science, which helps to extract information about a viewer's emotions by doing so-called "cold reads" (See Figure 1). We use an algorithm package called FaceReader (Nodules, 2014), where the six universal emotions, as well as contempt and neutral state, can be captured from recorded video of individuals' facial expressions. The demonstrated accuracy of this approach exceeds 90% (Langner et al., 2010).

The algorithm relies on three procedures. First, finding the face in an image where the position of the face is found using the Viola ones cascaded classifier algorithm (Viola and Jones, 2004). Second, a model-based method, the Active Appearance Model (AAM) (Cootes et al., 2004), synthesizes an artificial face model. It describes the location of 500 key points in the face and the facial texture of the area entangled by these points. The model uses a database of annotated images and calculates the main sources of variation found in the images. Principal Component Analysis is used to reduce the model dimensionality. Third, face classification the facial expressions is done by training an artificial neural network (Bishop, 1995). The network is trained to classify the six basic or universal emotions: happy, sad, angry, surprised, scared, disgusted as well as contempt and a neutral state. In addition, the dimensional emotions are also measured. Remarkably, the algorithm can also analyze a set of 20 action units of the Facial Action Coding System (FACS), such as cheek raising, nose wrinkling, dimpling, and lip tightening, to enhance the accuracy of emotion recognition (Langner et al., 2010). This technology has three advantages:

- *Unobtrusive.* FaceReader enables the measurement of emotions in a completely unobtrusive manner, eliminating the need for direct elicitation of data from research subjects. Participants are not physically contacted, and no additional devices are required. This unobtrusive measure mitigates the mere measurement effects that arise when participants know they are being measured.

- *Moment-to-moment.* The algorithm offers a real-time representation of categorical emotions (six universal emotions, contempt, and neutral state). It is worth noting that it carries out moment-to-moment measures of emotions at 30 frames per second, in other words, it can identify emotional changes in participants every 0.033 seconds.

- *Continuous calibration.* The system continuously re-calibrates to identify someone's neutral expression. It accounts for biases that might lead some individuals to appear consistently angry or happy due to their natural facial features. For example, certain facial characteristics, like narrowed lips and downturned corners of the mouth, may incorrectly suggest anger when that's actually part of the individual's neutral face. In our study, we treat neutral expression as a control variable to mitigate these individual-specific biases towards particular facial expressions. Moreover, if environmental factors hinder system improvements due to poor lighting conditions, continuous calibration helps counterbalance the bias. The algorithm continuously averages the facial expression intensities in the analysis in the following way: $ExpressionIntensity = max(0, (I_a - I_m)/(1 - I_a))$, where I_a is the expression intensity in the current frame and I_m is the average expression intensity over all frames before the current frame. If $(I_a - I_m)/(1 - I_a)$ is lower than zero, the calculated expression intensity will be zero. The intensity of Neutral is calculated in the analysis in the following way: $IntensityNeutral = (N_a + (1 - I_{max}))/2$, where N_a is the intensity of Neutral in the current

	Ad title	Brand	(Channel) Catergory	Matched Videos
Ad1	Misty Copeland Telling the best story: I WILL WHAT I WANT	Under Amour	Sports	(1) Champions Revealed: 2014 San Antonio Spurs- On Spurs Organization being family. (2) Tonight Show Superlatives: 2015 Super Bowl (3)Texas Volleyball: Pride, Passion & Tradition
Ad2	Best Nike Find Your Greatness Commercial: The Jogger	Nike		
Ad3	Foot Locker's Week Of Greatness: 2013 All Is Right	Foot Locker		
Ad4	He Got the Whole World In His Hands	Adidas		
Ad5	Ellen DeGeneres Super Bowl Commercial for Beats: The Right Music	Beats	Music	(1) Billboard THE HOT 100: Maroon "Sugar" (2) Taylor Swift : Award For Excellence - AMA Awards 2014 (3) The Voice 2014 Finale: Craig Wayne Boyd Original Performance: "My Baby's Got a Smile on Her Face"
Ad6	First iPod Commercial	iPod		
Ad7	Spotify Premium: Power to Music Lovers	Spotify		
Ad8	Philips Bluetooth: The Naked Funny Boy	Philips		
Ad9	TWC Super Bowl Commercial: The Walking Dead	Time Warner Cable	TV Show	(1) Game of Thrones Season 5 (2) The Big Bang Theory funny clip (3) House of Cards - Introductions
Ad10	DirecTV: Car Chase Commericial	Direct TV		
Ad11	Netflix Ad: Airport	Netflix		
Ad12	Best Apple TV Ad: Holiday	Apple TV		
Ad13	Little Baby's Ice Cream:This is a Special Time	Little Baby Ice Cream	Food (Cooking)	(1) Bizarre Foods America: Andrew's Austin Vlog (2) Man vs Food: Making The Absolutely Ridiculous Burger (3) Food Network: Drive-In Diners - Austin TX
Ad14	Carl's Jr. Super Bowl Ad: All Natural Burger	Carl's Jr		
Ad15	Snickers: The Brady Bunch	Snickers		
Ad16	Oreo Commercial: Bringing People Together	Oreo		

Table 1: Descriptions of Ads and Videos

frame and I_{max} is the maximum average intensity of all emotions in all the frames before the current one (Nodules, 2014).

3.2 Experiment and Design

3.2.1 Participants and Procedure

A total of 199 undergraduate students (mean age 21, age range 19 to 40, 42% female) participated in this study for partial course credit. Our experiment mimics the situation when people watch videos on YouTube in which each online video is preceded by an in-stream commercial. After the first few seconds of forced exposure (from 0s to 15s) consumers can skip away from a commercial to the video by clicking the "Skip Ad" button. The study featured 16 ads representing four different categories and brands. In parallel, 12 YouTube videos were selected from the four most popular channels. Each video's channel category was matched to one of the ad categories. For instance, *Tonight Show Superlatives: 2015 Super Bowl* matches the *Nike commercial* in the category of sports; *The Maroon's Sugar Video* corresponds to the *Beats ad* in the category of music. (See more details in Table 1.)

Each participant was exposed to a sequence of five ad-video sessions. In the beginning, each participant was asked to choose his/her 5 most favorite videos from the 16 videos on the list. We randomized the order of the 5 videos' presentations. In each session, an ad was randomly drawn from two sets of ads with the same probability and shown subsequently. One set contains the ads that are consistent with the video in the category and the other contains the complimentary set. Participants were told that they could skip the ad at any time after forced ad exposure by clicking the "Skip Ad" button similar to the YouTube interface. After fully or partially viewing the ad, participants were then exposed to the video. Thus, we randomized both the order of the ad and the video, eliminating biases caused by order effects. Before and after all five sessions were shown, participants were required to answer a number of questions related to each ad, such as ad familiarity, category interest, pre- and post-brand interest, a number of questions related to pre- and post-purchase intention as well as the demographics questions. Please see the exact wording of the questions in Table 2. To avoid the mere measurement effect, participants were not informed that their facial expressions were monitored until the debriefing form, which appeared at the end of the experiment. Upon acceptance, their facial expressions while watching the commercials were used in our study. The entire study spanned approximately 20 to 30 minutes.

3.2.2 Measurement

Ad-Skipping behavior. The main dependent variables are the participant's decision to skip and the duration of watching the ad. Similar in spirit to Teixeira et al. (2014); Texeira et al. (2012), we measure whether a participant chose to fully watch a particular ad or pressed the skip button prematurely. Then, we recorded the duration of the subject watching the commercial.

Emotion intensity. The main independent variables are represented by the intensity of a participants' emotion while watching the ads. Video material containing a participant's facial expressions was collected by means of a webcam installed on the laptop on which the participant watched the ad. After processing facial videos frame by frame with continuous calibration using AE technology (i.e. FaceReader), we obtain scores over time (around 30 measurements per second) representing a participant's emotions in real-time.

Control variables. We measure cognitive factors such as ad familiarity, category-specific interest, pre-brand specific familiarity, pre-/post-brand specific interest, pre-/post-purchase intention, and category consistency since prior research has shown these factors to influence attention (Teixeira

Measure	Time	Questions	Type
Individual-specific (Demographics)			
Age	Before	What is your age?	List of ages to choose.
Gender	Before	What is your gender?	Binary. "Female" or "Male"
Major	Before	What is your major?	List of majors to choose.
Catergory-specific			
Pre-category Interest	Before	How INTERESTED are you in the following categories in the ads?	5-point scale, anchored by "not at all interested" to "very interested"
Ad-specific			
Pre-brand familiarity	Before	How FAMILIAR are you in the following brands?	5-point scale, anchored by "not at all familiar" to "extremely familiar"
Pre-brand Interest	Before	How INTERESTED are you in the following brands?	5-point scale, anchored by "not at all interested" to "very interested"
Pre-purchase Intention	Before	How likely are you to PURCHASE the advertised brand?	5-point scale, anchored by "not at all interested" to "very interested"
Ad Familiarity	After	Have you EVER watched the ad before?	Binary, "Yes" or "No"
Post-brand Interest	After	How INTERESTED are you in the following brands in the ads?	11-point scale, anchored by "not at all interested" to "very interested"
Post-purchase Intention	After	How likely are you to PURCHASE the advertised brand?	11-point scale, anchored by "not at all interested" to "very interested"
Ad-Video consistency	During	Indicator function. "1"if category of ad matches movie trailer, "0" if not	

(Note: Column time indicates the time measured before/ during/ after the ad))

Table 2: Questionnaire Measurements and Format of Collection

et al., 2010). It is worth noting that we measured a novel control variable, the Ad-Video consistent indicator function, which is equal to "1" if the category of the ad matched the movie trailer, and "0" otherwise. We included this variable as a cognitive factor in our model to test whether viewers tend to watch an ad longer if the category of the ad is consistent with the category of the video. Furthermore, demographic information including age, gender, and major were also collected.

Ad frame-level variables. Leveraging state-of-the-art computer vision techniques, we extract two frame-level features from video advertisements, i.e. canny edge and foreground segmentation, as ad-specific time-variant control variables. First, existing research has shown that the visual complexity of advertising plays a central role in engaging viewers (Pieters et al., 2010). The Canny edge detection is a multi-step algorithm that can detect edges in the picture. We implement the canny edge algorithm to calculate the number of canny edges as an indicator of the visual complexity of the video frames. Second, we generate a foreground region (i.e. pixel-level mask for central object-like regions) by implementing a pre-trained deep fully convolutional neural network (Jain et al., 2017). This foreground can identify the key object in the video, and we quantify the area of the foreground object to measure the degree of visual distraction (See Figure 2).

Figure 2: Example: Region of Foreground Object in Advertisement 16

4 Model

4.1 Model Development

We propose a Bayesian dynamic generalized linear model (DGLM) to quantify the dynamic relationships between forced ad exposure, consumer emotions and ad-skipping behavior in the setting of online video advertisements. Previous work in psychology evidence and marketing considers a state-based approach to model emotions. Jeong et al. (1998) quantify and determine the way in which the emotional response to a video is reflected in the electrical activities of the brain, which is governed by some specific stochastic process. State-space models are largely employed to capture how the different valence of word-of-mouth evolve in the context of social media marketing (Godes and Mayzlin, 2004; Sonnier et al., 2011). Texeira et al. (2012) use a state-based duration model to specify the stock of consumer emotions.

We model skipping propensity in a way similar to existing work on advertising and awareness formation (Dubé et al., 2005; Bruce, 2008; Naik et al., 2008; Braun and Moe, 2013). In our model, the propensity to skip is modeled as a function of the stock of consumer emotions. We also consider how forced ad exposure impacts emotion. Based on the Nerlove-Arrow (NA) model we consider the evolution of skipping propensity as follows:

$$\frac{d\tilde{y}}{dt} = \beta' e_t + \varphi - \lambda\tilde{y}, \tag{4.1}$$

where $\tilde{y_t}$ denotes the skipping propensity and $\frac{d\tilde{y}}{dt}$ denotes the change in skipping propensity over time t. The vector e_t is a vector of emotion measures at time t. β denotes the effect of emotions on skip propensity, φ denotes the time-invariant intercept and λ represents the forgetting rate. In order to quantify the effects, we rewrite the model in discrete time as

$$\tilde{y_t} = (1-\lambda)\tilde{y_{t-1}} + \beta' e_t + \varphi + w_t, \tag{4.2}$$

where the error term w_t follows $N(0,1)$.

It is worth noting that we can not observe the true skipping decision during periods of forced ad exposure. Howver, the structure of state space models is such that missing observations can be easily accommodated in the filtering recursion (Petris et al., 2009). No additional adjustment is required in the smoothing recursion.

Our Bayesian DGLM model is formulated as follows. For viewer i=1,...I who is watching ad $j = 1, ..J$, at time $t = 0, ...T$, we denote a binary variable $Skip_{ijt}$ as:

$$Skip_{ijt} = \begin{cases} 1 & if \;\; d_{ijt} = 0 \;\; \& \;\; y^*_{ijt} > 0, \\ 0 & if \;\; otherwise. \end{cases} \tag{4.3}$$

$Skip_{ijt} = 1$ indicates viewer i skips in duration period t for ad j, otherwise $Skip_{ijt} = 0$. y^*_{ijt} denotes the latent variable skipping propensity, and d_{ijt} denotes the dummy variable that indicates forced ad exposure; d_{ijt} =1 indicates forced exposure for viewer i in duration t for ad j, else $d_{ijt} = 0$. The observation equations (4.4) and system equations (4.5) are formulated as:

$$\begin{pmatrix} y^*_{ijt} \\ e_{ijt} \end{pmatrix} = \begin{pmatrix} \tilde{y}_{ijt} \\ \tilde{e}_{ijt} \end{pmatrix} + \begin{pmatrix} v^y_{ijt} \\ v^e_{ijt} \end{pmatrix} \quad \begin{matrix} v^y_{ijt} \sim N(0,1) \\ v^e_{ijt} \sim N(0,V) \end{matrix} \tag{4.4}$$

$$
\begin{pmatrix} \tilde{y}_{ijt} \\ \widetilde{e}_{ijt} \end{pmatrix} = \begin{pmatrix} \rho & \beta' \\ 0 & \Delta \end{pmatrix} \begin{pmatrix} \tilde{y}_{ijt-1} \\ \widetilde{e}_{ijt-1} \end{pmatrix} + \begin{pmatrix} 0 \\ r^e \end{pmatrix} d_{ijt}
$$

$$
+ \begin{pmatrix} \eta' \\ 0 \end{pmatrix} z_{ij} + \begin{pmatrix} \phi' \\ 0 \end{pmatrix} a_{jt} + \begin{pmatrix} u_j \\ 0 \end{pmatrix} + \begin{pmatrix} w^y_{ijt} \\ w^e_{ijt} \end{pmatrix} \tag{4.5}
$$

$$
w^y_{ijt} \sim N(0, \sigma^2_w), \quad w^e_{ijt} \sim N(0, W)
$$

The vectors e_{ijt} and e^*_{ijt} are the observation and state vectors for emotion, respectively. The emotion intensities include *Happiness*, *Suprise*, *Sadness*, *Anger*, *Fear*, *Disgust*, *Contempt* and a *Neutral* state. The scalars y^*_{ijt} and $\tilde{y}_{ijt}$ are the observation and state variables for skipping propensity, respectively. The vector z_{ij} denotes the vector of individual time-invariant control variables $Ad\ Familarity$, $Pre-brand\ Interest$, $Pre-catergory\ Interest$, $Pre-purchase\ Intention$, *Category Consistent* (i.e. dummy variable indicates whether the category of the ad was matched to the channel category of the video). The vector a_{jt} denotes the vector of time-variant ad-specific control variables for each ad, including ad visual *Complexity* and *Distraction*. The scalar u_j represents the ad-specific fixed effect in the system equations of skipping propensity. The scalar v^y_{ijt} and vector v^e_{ijt} are observational errors for skipping propensity and emotion, respectively. The scalar w^y_{ijt} and vector w^e_{ijt} are system errors for skipping propensity and emotion, respectively. All the error terms are assumed to be independent. The scalar ρ and matrix Δ are parameters that measures the decay in state variables over time. The vectors β, η, and ϕ measures the effect of emotions, individual time invariant controls and ad specific time varying controls, respectively, on ad-skipping behavior. The vector r^e measures the effect of forced ad exposure on emotions.

4.2 Model Estimation

We use Markov Chain Monte Carlo (MCMC)To estimate the parameters of our Bayesian DGLM. We estimate the model in R. To expedite the estimation process, we integrate C++ code into R through Rcpp packages. We specify normal priors on [ρ, β, Δ, r^e, η, ϕ u], the parameters of the observation and system equations. We specifiy an Inverse Gamma prior on the state variance of skipping propensity, σ_w^2. We specify and Inverse Wishart prior on the and covariance matrices of emotion observation and state equations, V and W, respectively. We use the forward-filtering and backward-smoothing algorithm for skip propensity and emotion state variables $\widetilde{y}_{ijt}$ and $\tilde{e_{ijt}}$, respectively (West and Harrison, 1997). The latent y^*_{ijt} is generated by a standard binary probit data augmentation step. Conditional on the state variables $\widetilde{y}_{ijt}$ and $\tilde{e_{ijt}}$ as well as the augmented y^*_{ijt} all other parameters in the model are estimated through a Gibbs sampler. More details on the priors, conditional posteriors, and estimation can be found in Appendix A.

To test our estimation procedure we construct a simulated data set using the proposed model and a set of known parameters. In Appendix B we show that the model ably recovers the data generating parameters. Of particular importance is the models ability to recover unobserved intended skipping behavior. We can not observe the intended skipping decision during the forced watching period. Therefore, $Skip_{ijt|d_t=1}$ are considered as missing observations. To identify the parameters in the skipping propensity state equation, we design the forced ad exposure condition in the experiment. For each viewing instance, each participant is forced to watch the first 0s, 5s, or a random draw from {6s,7s,...15s}, with the same probability of each forced exposure condition. Therefore, the data set consists of the 1/3 no missing samples and 2/3 partially missing samples. We employ a Bayesian approach to draw random samples of missing data based on information contained in the observed data. In Appendix B we show that by imputation from the distribution $P(\Theta, Skip^{mis}|Skip^{obs})$ (where Θ denotes all the parameters and $Skip^{mis}$, $Skip^{obs}$denote observed and missing data on skipping), we can reliably identify intended skipping behavior in the forced exposure condition.

5 Data and Results

5.1 Data Description

In our experiment, we collected 995 viewing instances from our 199 participants (See Summary Statistics in Table 3). To the best of our knowledge, our study is the first to measure emotions in a completely unobtrusive way. However, a cost of unobstrusiveness is that the FaceReader algorithm was unable to calibrate 50% of these viewing instances. This is generally due to poor light conditions, unrecognizable pose orientation, image with poor contrast, the image of the face obstructed by unconscious gestures, or subjects wearing glasses, which hinders classification. To realize complete unobtrusiveness, missing data is inevitable, mainly because participants are not aware of being monitored. We assume that not including these data, which are missing at random, does not significantly bias our results as these unobserved responses are unlikely to correlate with our dependent variable. After eliminating the missing data, there are 480 viewing instances with 1,215,870 observations on specific time points. As an illustrative example, histograms of watching duration for three ads are shown in Figure 3 (Top row). Ad1 is a commercial for apparel maker Under Armour, Ad5 is a commercial for Beats's headphones. Ad 14 is a Super Bowl commercial for fast food restaurant Carl's Jr.

Due to the aforementioned challenges that arise when measuring emotions unobtrusively, each of our 480 viewing instances contains some points with missing emotion intensity data. In total about 3% of our approximately 1.2 million observations contain moments of missing data. To handle the missing data of emotion intensity in each viewing instance, we consider a number of smoothing techniques, including Average Smoothing, Lagrange Interpolating Polynomial, Newton Interpolating Polynomial, and Cubic Spline Smoothing. Based on our results, the Cubic Spline smoothed the emotion curve with the highest accuracy due to the advantage of using all of the available data to construct a cubic between each pair of continuous points with continuous first and second derivatives. To reduce the computational burden in estimating more than 1,215,870 latent states, we implemented the exponential moving average (EMA) filter to smooth the granular emotion data and then shrink the data to 2 fps (frame per second). By weighting each observation with the most recent emotion changes, EMA can filter out significantly high-frequency noise and retain the most information after data shrinkage. After data pre-processing, the time series plots of average emotions over all the respondents for our three example ads are shown in Figure 3 (Bottom row).

Upon visual examination of the duration histograms and time series plots of average emotions, we can see some model-free evidence in Figure 3. Time series plots reveal some potential evidence of a possible relationship between forced ad exposure and emotions. In the bottom panel of Figure 3 , for example, happy emotions sharply decline while disgust has higher intensity during the first few seconds in Ad1. Angry emotion dominates at the beginning of Ad 5, and Ad 14. We can also clearly see that the interested viewers do not necessarily watch the ad for a longer time but more viewers skip the ad at the same time when the surprise sharply goes down. While this model-free

	Mean	Standard Deviation	Min	Max
Ad views	Watching Duration(s)			(Ad Length)
Ad1	42.02	21.22	10.0	60
Ad2	45.89	21.12	9.0	62
Ad3	37.57	23.74	4.5	60
Ad4	51.43	15.96	6.0	60
Ad5	45.10	23.39	4.0	67
Ad6	36.34	22.26	3.0	60
Ad7	43.90	21.17	6.5	60
Ad8	37.86	23.24	6.5	63
Ad9	41.80	19.47	12.0	60
Ad10	40.63	21.90	7.0	62
Ad11	42.81	21.85	9.5	60
Ad12	57.68	37.69	2.0	90
Ad13	40.41	23.14	5.0	66
Ad14	36.87	16.71	7.0	52
Ad15	29.56	19.16	6.0	54
Ad16	47.77	23.79	2.5	67
Emotions	Intensity			
Happiness	0.013	0.083	0.00	1.00
Sadness	0.012	0.075	0.00	1.00
Anger	0.033	0.132	0.00	1.00
Surprise	0.040	0.152	0.00	1.00
Fear	0.007	0.047	0.00	1.00
Disgust	0.022	0.090	0.00	1.00
Contempt	0.013	0.053	0.00	1.00
Valence	-1.022	102.267	-14930.00	5944.00
Arousal	0.120	3.155	-490.70	10.91
Neutral	0.254	0.326	0.00	1.00
Forced Ad Exposure	5.00	4.64	0	15
Control Variables				
Individual-specific				
Gender	0.49	0.50	0	1
Catergory-specific				
Pre-category Interest	2.68	1.37	1	5
Ad-specific				
Pre-brand familiarity	2.24	1.37	1	5
Pre-brand Interest	2.65	1.49	1	5
Pre-purchase Intention	3.11	1.53	1	5
Ad Familiarity	0.17	0.37	0	1
Post-brand Interest	4.52	3.50	0	10
Post-purchase Intention	3.93	3.45	0	10
Ad-Video consistency	0.49	0.50	0	1

Table 3: Descriptive Statistics of Ad Views, Emotions and Control Variables

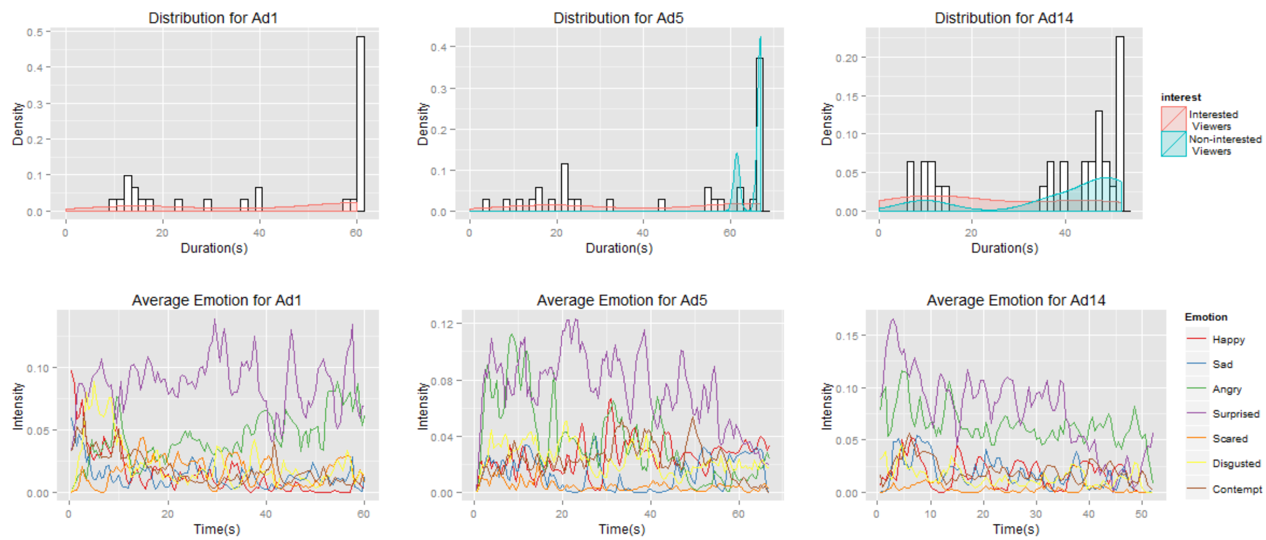


Figure 3: **Data Summary. Top row:** Histograms of watching duration over all the respondents for Ad1, Ad5, Ad14. **Bottom row:** Time series plots of emotions over all the respondents for Ad1, Ad5, Ad14.

evidence is promising, we need to get down to the individual level because the length of forced ad exposure and watching duration varies across individuals. In the next Section, we employ a model-based individual-level approach to investigate the effect of forced ad exposure and the resulting emotions on ad-skipping decisions.

5.2 Estimation Results

5.2.1 *Does forced ad exposure affect emotions?*

Existing literature has demonstrated the negative outcomes stimulated by forced exposure. Forced ad exposure leads to a negative perception of the advertising as intrusive and elicits the viewer's involuntary attention, making the viewer attempt to restore their freedom by skipping (Edwards et al., 2002). We conceptualize the role of forced ad exposure in ad-skipping behavior and demonstrate how the effect may be moderated by emotions. To analyze the effect of forced ad exposure on emotions, we run the model based on categorical emotion (i.e. six universal emotions).

Table 4 summarizes the parameter estimates for the emotion state equation. All the parameters ($\Delta_{Happiness}$, $\Delta_{Sadness}$, Δ_{Anger}, $\Delta_{Surprise}$, Δ_{Fear}, $\Delta_{Disgust}$, $\Delta_{Contempt}$) in the state matrix are significantly positive, indicating the high dependence on the previous emotion state. Anger,

Parameter	Posterior Mean	Posterior SD	Posterior 95% CI
$r^e_{Happiness}$	0.134*	0.042	(0.052,0.220)
$r^e_{Sadness}$	0.022	0.012	(-0.002,0.044)
r^e_{Anger}	0.062	0.153	(-0.236,0.350)
$r^e_{Surprise}$	-13.820*	2.513	(-18.070,-9.854)
r^e_{Fear}	-0.063	0.087	(-0.209,0.114)
$r^e_{Disgust}$	-0.929*	0.471	(-1.759,-0.118)
$r^e_{Contempt}$	4.283*	0.812	(2.999,5.613)
$\Delta_{Happiness}$	0.474*	0.003	(0.469,0.478)
$\Delta_{Sadness}$	0.624*	0.001	(0.622,0.625)
Δ_{Anger}	1.185*	0.050	(1.127,1.288)
$\Delta_{Surprise}$	0.866*	0.024	(0.826,0.906)
Δ_{Fear}	0.855*	0.001	(0.853,0.858)
$\Delta_{Disgust}$	0.543*	0.005	(0.533,0.549)
$\Delta_{Contempt}$	0.851*	0.024	(0.812,0.891)

*95% credible interval does not span zero

Table 4: **Parameter Estimates for Emotion State Equation:** Effect of Forced Ad Exposure on Categorical Emotions and Emotion State Dependency

surprised, and scared emotions show a high correlation, suggesting that the underlying change of those emotions largely depends on their previous emotional states. Notably, forced ad exposure does have a significant impact on the underlying change of four emotions, happiness, surprise, disgust, and contempt. Specifically, forced ad exposure statistically decreases surprise (with negative coefficient $r_{Surprise} = -13.820$), and disgust (with negative coefficient $r_{Disgust} = -0.929$). In contrast, contempt (with positive coefficient $r_{Contempt} = 4.283$) significantly increases when forced ad exposure is present. The results are in accordance with the literature reviewed. It is not surprising that forced ad exposure evokes negative feelings, such as contempt. Our results also emphasize the suppression of positive emotions, which is rarely explored in existing literature. For instance, the greater value of the coefficient on surprise, -13.820, indicates the strong and detrimental power that forced exposure can suppress the viewer's surprised feeling. In addition, the effect of the forced ad exposure on emotion relies on how emotions are classified. The broad classification of the dimensional model cannot capture this effect.

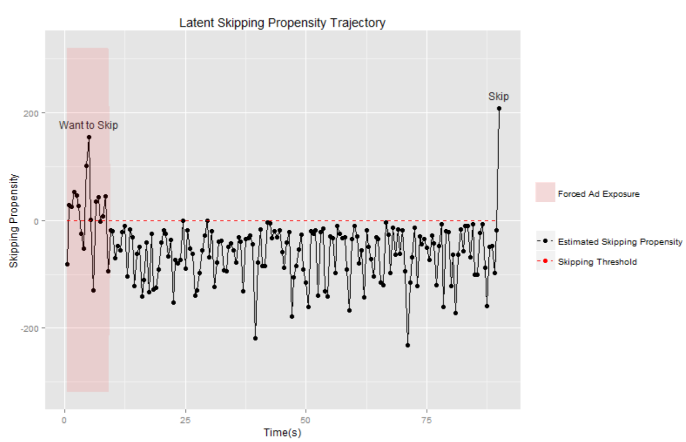
5.2.2 *What are the effects of emotions on ad-skipping behavior?*

We investigate how the resulting emotions can influence ad-skipping decisions. Due to the abundant literature on how to dynamically transform perceived categorical emotions to advertising technologies (Elpers et al., 2004), we establish more in-depth substantial implications using the category emotions model (six universal emotions). Table 5 reports the parameter estimates for skipping propensity state equation. As we can see, the effect of cognitive factors such as ad familiarity, ad pre-category interest, pre-brand interest, pre-purchase intention, and ad-video as well as the demographic information, gender, are not significant. In contrast, categorical emotions have a significant effect on ad-skipping behavior. Our data show that ad viewing duration may be mostly about emotions, highlighting a major limitation of extant research where only cognitive factors are identified as affecting ad-skipping behavior. Table 5 also reports four significant emotions. Surprise (with negative coefficient $\beta_{Surprised} = -0.940$) causes a conditionally significant increase in the skipping probability and anger (with negative coefficient $\beta_{Anger} = -14.990$) causes a statistically significant decrease in skipping probability. In contrast, sadness (with positive coefficient $\beta_{Sadness} = 4.702$), disgust (with positive coefficient $\beta_{Disgust} = 3.756$) and contempt (with positive coefficient $\beta_{Contempt} = 9.684$) significantly increase the risk of losing audience's attention. The results are also consistent with existing behavioral research that shows most advertisers utilize humor, which arouses emotions of surprise to engage viewers (Elpers et al., 2004; Texeira et al., 2012). We extend their results to multiple dimensions of emotions, especially negative emotions. Counter-intuitively, the contempt and disgust drive viewers to skip away while angry feelings contribute to engaging them. In addition, visual complexity, which is uncorrelated with emotion variables plays a conditionally significant role in increasing skipping-propensity. It is interesting to note that the decay factor ($\rho = 0.790$) indicates a high carryover effect on skipping propensity. The results confirm that viewers' emotions at the current time period can largely affect their future ad-skipping decisions. Therefore, forced ad exposure provides an opportunity for advertisers to set the right tone and engage their audience in the opening seconds.

Parameter	Posterior Mean	Posterior SD	Posterior 95% CI
ρ	0.790	0.003	(0.783,0.796)
$\beta_{Happiness}$	-0.072	3.580	(-6.922,6.774)
$\beta_{Sadness}$	4.702*	1.990	(0.690,8.758)
β_{Anger}	-14.990*	4.511	(-23.580,-6.526)
$\beta_{Surprise}$	-0.940†	0.747	(-2.446,1.443)
β_{Fear}	-2.592	1.439	(-5.472,0.168)
$\beta_{Disgust}$	3.756*	1.286	(1.107,6.362)
$\beta_{Contempt}$	9.684*	2.310	(5.180,14.400)
η_1(Ad Familiarity)	0.146	0.308	(-1.910,1.851)
η_2(Pre-category Interest)	-0.587	0.606	(-1.144,1.194)
η_3(Pre-brand familiarity)	-0.596	0.661	(-1.272,1.349)
η_4(Pre-brand Interest)	-0.406	0.648	(-1.223,1.264)
η_5(Pre-purchase Intention)	0.701	0.596	(-1.190,1.068)
η_6(Ad-Video consistency)	-1.113	0.902	(-1.777,1.746)
η_7(Gender)	-0.685	0.873	(-1.907,1.622)
$\phi_{Complexity}$	-0.001	0.021	(-0.042,0.040)
$\phi_{Distraction}$	0.004*	0.002	(0.000,0.008)

*95% credible interval does not span zero; †marginally significant

Table 5: **Parameter Estimates for skipping Propensity State Equation:** Effect of Categorical Emotions and Cognitive Factors on Ad-Skipping Decision and Carryover Effect



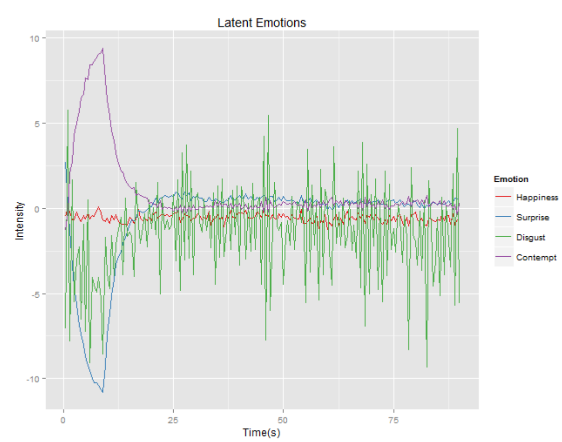


Figure 4: Estimated Skipping Propensity Trajectory of y^*_{ijt} and Estimated Latent Emotions $\tilde{e_{ijt}}$ for one viewing instance

Forced Ad Exposure	% of Viewers Watch more than 30s	%of Viewers Watch Entire Ad	Average Duration	Proportion of Average Voluntary Duration
0s	0%	0%	3.04s	0.05
5s	30%	20%	15.42s	0.19
10s	32%	32%	36.24s	0.54
15s	40%	38%	38.29s	0.52
20s	38%	38%	38.97s	0.05
25s	34%	34%	39.57s	0.39

Table 6: "What if" Analysis

Notes: % of Viewers Watch more than 30 seconds represents the retention rate of viewers past the 30-second mark, indicating the ad's ability to not only retain audience attention but also generate advertising revenue; % of Viewers watch the entire ad represents the retention rate of viewers who watch towards the end of the ad; Average Duration provides the average length of the watching duration shown to viewers; Proportion of Average Voluntary Duration shows the ratio of the average length of watching duration to length of non-forced watching period. A higher proportion indicates a larger voluntary engagement.

5.2.3 *What is the most effective forced watching time?*

Given the results on the effect of forced ad exposure on emotions and the resulting emotions on ad-skipping decisions, we address questions on the optimal forced watching duration. We first discuss the pros and cons of forced ad exposure based on our results. To provide more intuition, Figure **??** presents the estimated emotions state variable and skipping propensity state variables for one viewing instance. On the one hand, forced ad exposure sharply suppresses the surprised emotion and ignites the contempt emotions, resulting in a higher skipping propensity. On the other hand, forced watching may provide potential benefits. Although the resulting skipping propensity exceeds the skipping threshold at some time periods, the forced watching requirement causes the impatient viewers to wait until the most exciting part of the ad. As we can see in Figure **??**, the level of surprise sharply goes up after the negative spike during forced exposure. Then the viewer watched the whole ad towards the end. This suggests that online platform marketers and designers pay attention to how to minimize the negative outcomes while taking advantage of the potential benefits of forced ad exposure.

To further illustrate how to balance the trade-off on forced ad exposure, we conduct a "what if" analysis based on our model. We compare the simulated distributions of viewing duration for Ad1 with different lengths of forced ad exposure. Through a series of simulations to quantify the impact of forced ad exposure, we find that while it does increase viewership, longer forced ad exposure doesn't necessarily guarantee higher viewership. It is important to note that these

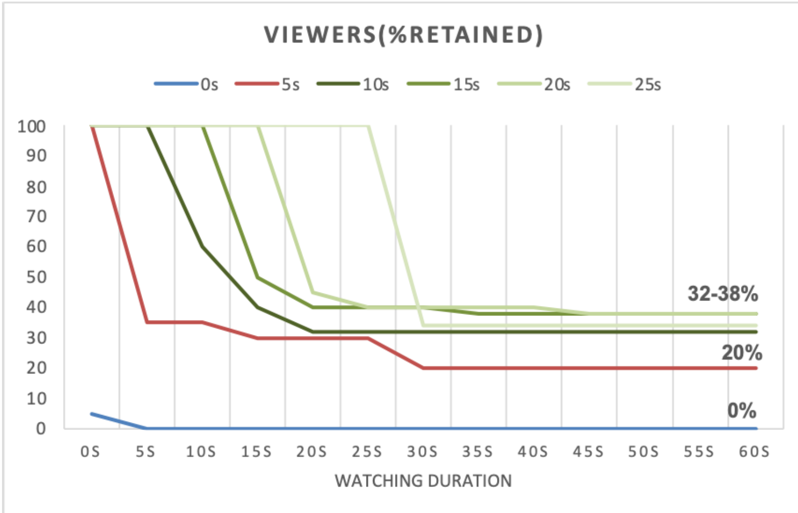


Figure 5: “What if” Analysis

changes in distribution significantly affect the revenue of advertisers and online platforms, especially considering the YouTube TrueView model, where advertisers pay only when at least 30 seconds or the entire ad is viewed. We measure the percentage of viewers watching more than 30 seconds to provide implications on the strategic pricing strategy between the online video platform and advertisers, depending on their different objective functions. For example, as shown in Table **??**, after increasing the forced watching time from 0 to 10 seconds, 32% of total respondents surpass the 30-second mark and continue watching the ad. Conversely, 6% of viewers drop before the 30-second mark when the forced viewing time is extended from 15 to 25 seconds. Overall, a forced ad exposure duration of 10-15 seconds results in the highest viewership. When considering the average viewing duration, it is not surprising that longer forced exposure results in an extended duration. However, focusing on the average voluntary duration (i.e., duration during the non-forced time periods), the longer forced ad exposure may have a negative impact. Normalizing the average voluntary duration by dividing the length of non-forced watching periods, we find this proportion also peaks at 10-15 seconds. As depicted in Figure 5, the skip rate without forced ad exposure, using the immediately skippable strategy, is exceedingly high, almost 100%. We propose that extending the forced ad exposure from 5 seconds (the current YouTube strategy) to 10 seconds for this ad could lead to at least a 12% increase in viewers watching past the 30-second mark. It is important to note that in this simulation, we assume there is no non-transparent skipping behavior. Future research could compare simulated distributions while considering non-transparent skipping decisions.

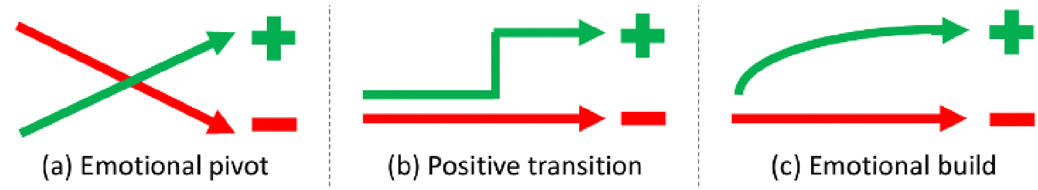


Figure 6: Three Archetypes of Dramatic Structure in Product Ads

5.2.4 *How to design the best emotional appeal ad?*

We have learned that there is a significant carry-over effect, and it is established that surprise (and anger) are considered "positive" emotions that effectively engage viewers. Conversely, contempt, sadness, and disgust are categorized as "negative" emotions that prompt viewers to decide to skip the content. In this subsection, we delve into an examination of the three common archetypes of dramatic structure in product ads, based on our parameter estimates. Additionally, we offer practical guidelines on how to craft the most effective emotional appeal in an advertisement. The framework outlined in the Advertising Research Handbook (O'Donohoe and Young, 2008) underscores the substantial connection between the structural composition of advertisements and the emotional tone they evoke within video content. This framework delineates two primary structures, as illustrated in Figure 6. The *emotional pivot* structure initiates with a negative emotional tone, gradually transitioning towards a heightened positive sentiment. Conversely, the *emotional build* entails a progressive escalation of positive sentiment leading to a climax. Thus, the emotional tone is as pivotal to comprehending the narrative of an ad video as the pivotal climax. Given that advertising is geared towards engaging an audience and persuading them to take action, the emotional response elicited from the viewer stands as the paramount element. We compare the simulated distributions of watching duration for Ad 1 with different types of emotion trajectories: (a) Emotional pivot, (b) Positive transition, and (c) Emotional build (Ye et al., 2018). Based on our estimated parameters, we find that the emotional build can retain the most viewers, due to the high carry-over effect. In order to quantify the ingredients that lead to the final emotional build and to offer suggestions for practitioners, we perform the follow-up video analysis for Ad1: (1) qualitative analysis of Freytag's pyramid and (2) quantitative analysis of the correlation between multi-modal features and emotions.

First, we revisit Freytag's pyramid and explore whether there is such a pattern in Ad 1. Freytag's pyramid is a well-known method often employed in storytelling, characterized by the following sequence: commencing with exposition, then advancing through rising action, reaching the pinnacle at the climax, and culminating in the denouement or resolution which depicts the declining action (Harun et al., 2013). Taking Ad1 as an example, we can see that the video pattern matches Freytag's pyramid, as shown in Figure 7. This suggests that Freytag's pyramid is indeed a creative narrative technique for ad designers or advertisers to reach the emotion climax for the emotional build.

Then, another important question for practitioners is how to actually design videos with the right set of features in order to evoke such an emotional response. To delve deeper into this, we take a multimodal approach to investigate the role of different types of sensory data in evoking viewers' emotions. The consumer experience is inherently multimodal, with vision and audio being the most paramount modalities. When viewing a video advertisement, we engage with both the visual frames (depicting individuals and their surroundings) and the auditory elements (such as music, voice-overs, sounds of human activities, etc.). These two modalities complement each other and significantly contribute to our overall perception and emotional response. Specifically, as shown in Figure , we extract the audio signal from the ad video and conduct the acoustic spectrum analysis. The Log Power Spectrogram illustrates the time-frequency representation of the one-minute audio. We observe more intense sound events and lower-frequency components—largely arising from the inspiring music when the Ballet dancer starts to dance—during the second half of the video. On the visual front, we utilize computer vision tools for visual motion analysis. Employing the Farneback method (Farnebäck, 2003) to process each set of consecutive video frames, we extract dense optical flow, capturing the object motion between these frames and the camera motion. The bottom right figure visualizes an example of the computed optical flow for two visual frames that capture the jumping action of the Ballet dancer, where the lighter color indicates larger motions. It accurately portrays the motion of the dancer in the video.

After extracting the per-moment acoustic and visual features, we aggregate them temporally to see whether there is any correlation with the viewer's emotions in the video. To do this, we compute the audio energy (the root mean square value for each audio frame) and the motion strength (the aggregated magnitude of optical flow for the entire visual frame) at each timestamp. Plotting these values across the entire time frame shown in Figure 9, we aim to identify patterns that correspond

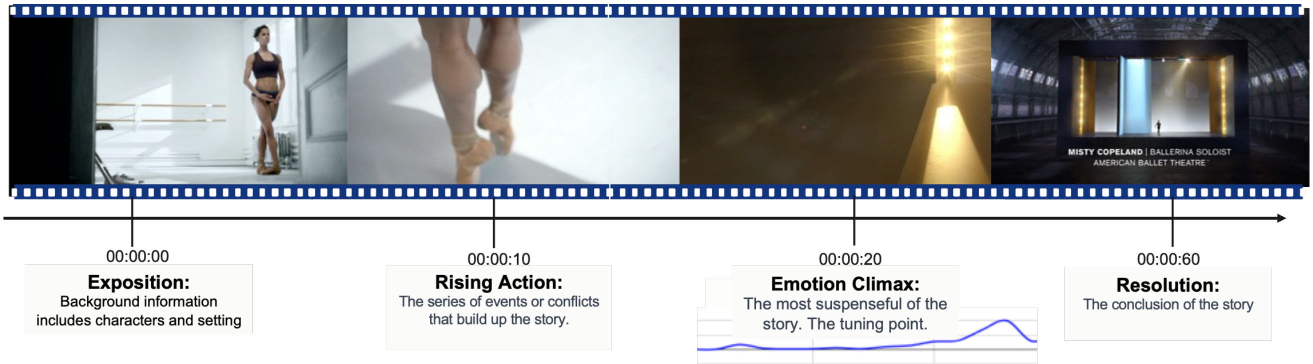


Figure 7: Freytag's Pyramid Structure in Ad1

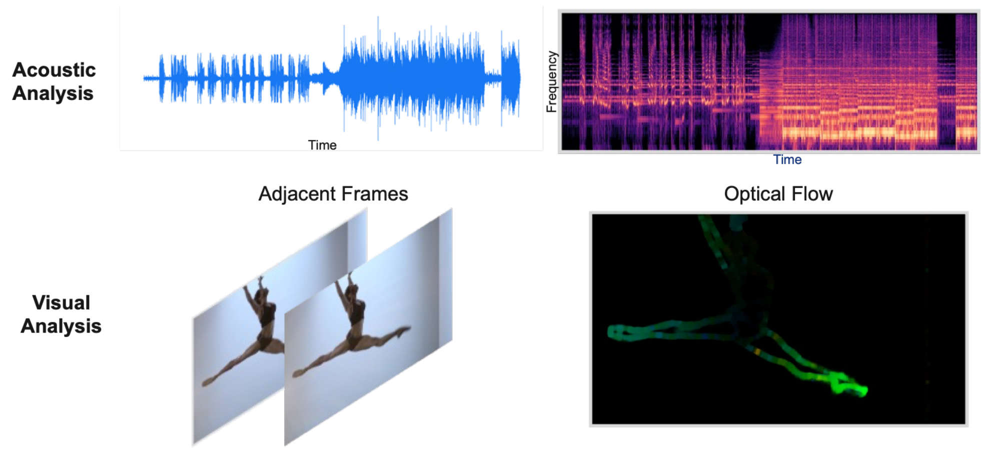


Figure 8: Acoustic and Visual Features

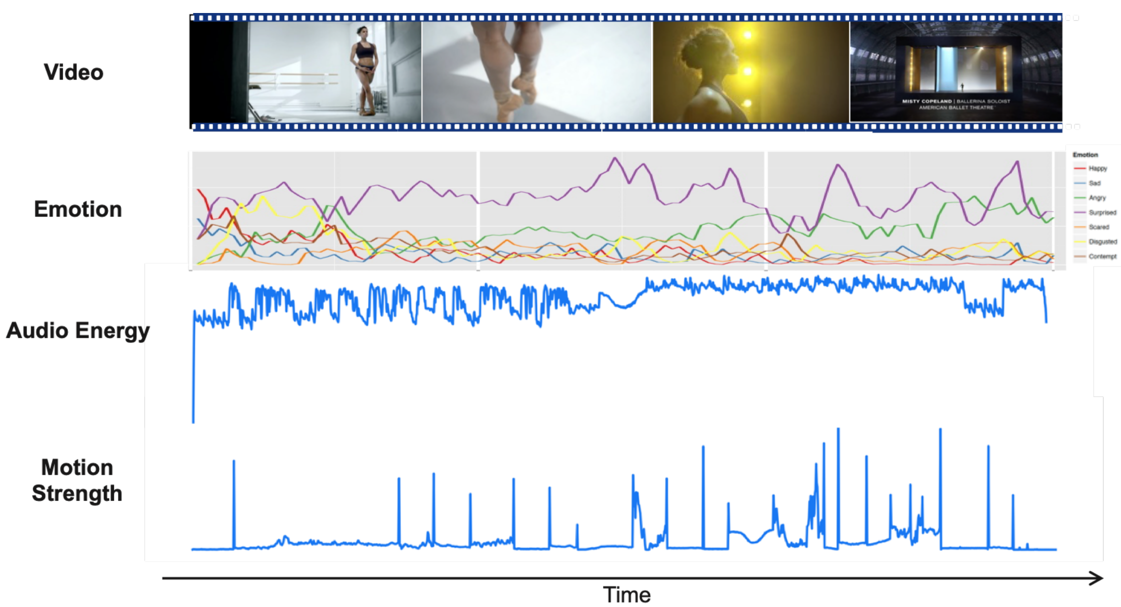


Figure 9: Acoustic, Visual Features and Emotions Trends

to the viewer's emotional journey. In our analysis, we discover that specific peaks in both audio energy and visual motion strength, which also coincide with particular emotional shifts in the video. These shifts are notably aligned with what could be identified as the climactic moments in the storyline. For instance, during the high-energy segments of the video, characterized by elevated audio energy and intensified visual motion, viewers exhibit increased emotional engagement, possibly corresponding to the climax or the most emotionally impactful scenes. This synchronization between the peaks in audio energy, visual motion, and the viewer's emotional response suggests a correlation between the audio-visual elements and the emotional impact on the audience, particularly during the climactic points of the video. This suggests that ad designs can intentionally adjust the corresponding visual and acoustic contents in the ads to achieve the desired emotional trajectory of the viewers.

6 Discussion

The ascent of online video advertisements on prominent social platforms like YouTube and LinkedIn is gaining significant momentum. This surge has prompted digitally astute brands to redirect their advertising budgets from traditional TV commercials to online video ads. According to ComScore (2013), YouTube garners over 1 billion views per day, signifying its enormity. If YouTube were a search engine, its scale would rank second, almost double the size of Bing and Yahoo combined, and about a third the size of Google.com.

Our results provide valuable implications for ad designers, advertisers, and online video platforms. Most prevalent online platforms allow viewers to skip or terminate advertisements. A pivotal and unresolved query for online networks remains whether to mandate forced ad exposure on a skippable ad and to what extent they should allow viewers the freedom to maximize viewership. Our proposed model provides insights into determining the most effective forced-watching time. Currently, viewership rates for YouTube 5-second pre-roll ads are generally lackluster, with few marketers customizing ads specifically for pre-roll. They prefer airing TV ads without alterations. The high carry-over effect in our findings underscores the significance of initial seconds. Advertisers failing to recognize the critical impact of these opening moments risk losing the audience, thereby failing to convey the ad's essential content.

Although there's no perfect formula for attention-grabbing ad design, our research quantifies the effect of forced ad exposure on emotions and identifies the emotional drivers behind ad-skipping decisions by measuring viewers' facial expressions. Our model is unique in that it conceptualizes the role of forced ad exposure on ad-skipping decision-making and demonstrates how emotions may mediate this effect. This area remains underexplored in previous literature. Furthermore, our paper pioneers a new research field focused on emotion/emoji targeting to harness moment-to-moment emotions for improved advertising effectiveness.

6.1 Implications for Ad Designers and Advertisers

Upon examining hundreds of reactions to numerous ads, second by second, and tracking precisely when viewers cease watching, we find that keeping viewers engaged is primarily tied to emotions, particularly surprise. In contrast, the desire to skip ads largely hinges on emotions like contempt and disgust. While surprise is considered a "positive" emotion that most designers seek to invoke, it is equally crucial to avert "negative" emotions such as disgust, contempt, and sadness, given their greater impact compared to surprise. Additionally, our findings reveal that anger boosts viewership in some special cases. Future ad designs might benefit from tactfully integrating mildly controversial content to increase viewership. The sustained carryover effect of skipping propensity underscores the importance of captivating viewers from the outset. While TV ads may have worked decades ago, today's online audiences require initial hooking moments. The dynamic nature of skipping propensity implies that engaging viewers from the start lowers the probability of subsequent skipping.

We also suggest that forced ad exposure does not fundamentally engage viewers but rather moderately delays ad-skipping decisions, offering designers an opportunity to retain viewers through emotion-induced techniques. Facial tracking technology has proven to be a valuable method for pre-testing advertising techniques by gauging viewers' perceived emotions. Additionally, the "emotional build", characterized by a gradual increase in positive sentiment leading to a climactic peak, stands out as the most effective strategy for nurturing positive emotions. Using Freytag's pyramid, a creative narrative technique, could be beneficial for ad designers and advertisers aiming to evoke an emotional climax. The correlation between multi-modal audio-visual elements and their emotional impact suggests the importance of aligning visual and acoustic content in ads to achieve the desired emotional trajectory in viewers.

6.2 Implications for Online Video Platforms

Our model provides insights on whether online video platforms should impose forced ad exposure on a skippable ad and what degree of freedom they should give to the viewers in terms of maximizing viewership/revenue. Although most video streaming companies are imposing forced ad exposure on a skippable ad, its negative effect cannot be ignored as it largely ignites contempt and disgust and suppresses the surprised and happy feeling. Moreover, within the YouTube TrueView model, the distribution of watching duration shapes the interaction between advertisers and online platforms in terms of design and pricing strategy. Google has introduced an in-house measurement system called Active View for assessing ad visibility. Incorporating a novel ad testing experiment via facial tracking technology could simulate potential viewing duration distributions, thereby offering more informative prior knowledge for pricing strategies.

6.3 Future Research

As a first step towards unveiling pre-roll skippable ads, this study inherently contains several limitations, paving the way for potential future research opportunities. First, the current model lacks consideration of heterogeneity. Future advancements could involve estimating individual-level effects and leveraging parallel computing to mitigate the substantial computational load currently faced. Second, expanding the model to accommodate time-variant effects is essential. Traditionally, advertisers construct narratives that build towards a dramatic climax or a surprising conclusion. Balancing the allocation of resources between enhancing the ad's beginning and strategically placing engaging elements towards the end poses a challenge for designers. Third, examining the disparities between pre- and post-cognitive factors offers an avenue for investigation into the effects of forced ad exposure on heightened awareness metrics such as brand interest and purchase intention. Even in cases where viewers react negatively to forced exposure ads, they are still subjected to the advertisement's message. Future research could delve deeper into the long-term impacts of forced ad exposure. Fourth, While the present focus of the paper centers on pre-roll ads, where a skippable ad precedes the video, future research can explore the intricate interactions between the ad content and the video itself (Fong et al., 2024). Last, this study represents an initial stride towards understanding multi-modal features, with further exploration encouraged, especially in the context of the emerging field of generative models.

References

Abelson, R. P. and Sermat, V. (1962). Multidimensional scaling of facial expressions. *Journal of Experimental Psychology*, 63(6):546.

Alden, D. L., Mukherjee, A., and Hoyer, W. D. (2000). The effects of incongruity, surprise and positive moderators on perceived humor in television advertising. *Journal of Advertising*, 29(2):1–15.

Aravindakshan, A. and Naik, P. A. (2011). How does awareness evolve when advertising stops? the role of memory. *Marketing Letters*, 22(3):315–326.

Bishop, C. M. (1995). *Neural networks for pattern recognition.* Oxford university press.

Braun, M. and Moe, W. W. (2013). Online display advertising: Modeling the effects of multiple creatives and individual impression histories. *Marketing Science*, 32(5):753–767.

Bruce, N. I. (2008). Pooling and dynamic forgetting effects in multitheme advertising: Tracking the advertising sales relationship with particle filters. *Marketing Science*, 27(4):659–673.

Chang, H. H. and Pham, M. T. (2013). Affect as a decision-making system of the present. *Journal of consumer research*, 40(1):42–63.

ComScore (2013). Comscore releases june 2013 u.s. online video rankings.

Cootes, T. F., Taylor, C. J., et al. (2004). Statistical models of appearance for computer vision.

Dubé, J.-P., Hitsch, G. J., and Manchanda, P. (2005). An empirical model of advertising dynamics. *Quantitative marketing and economics*, 3(2):107–144.

Dukes, A. and Liu, Q. (2023). The consumption of advertising in the digital age: Attention and ad content. *Management Science.*

Dukes, A. J., Liu, Q., and Shuai, J. (2019). Skippable ads: Interactive advertising on digital media platforms. *Available at SSRN 3169629*.

Edwards, S. M., Li, H., and Lee, J.-H. (2002). Forced exposure and psychological reactance: Antecedents and consequences of the perceived intrusiveness of pop-up ads. *Journal of Advertising*, 31(3):83–95.

Ekman, P. and Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica*, 1(1):49–98.

Ekman, P., Friesen, W. V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., Krause, R., LeCompte, W. A., Pitcairn, T., Ricci-Bitti, P. E., et al. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of personality and social psychology*, 53(4):712.

Elpers, J. L. W., Mukherjee, A., and Hoyer, W. D. (2004). Humor in television advertising: A moment-to-moment analysis. *Journal of Consumer Research*, 31(3):592–598.

Farnebäck, G. (2003). Two-frame motion estimation based on polynomial expansion. In *Image Analysis: 13th Scandinavian Conference, SCIA 2003 Halmstad, Sweden, June 29–July 2, 2003 Proceedings 13*, pages 363–370. Springer.

FinancialTimes (2015). Youtube to provide viewer data to advertisers.

Fong, H., Kumar, V., and Sudhir, K. (2024). A theory-based explainable deep learning architecture for music emotion. *Marketing Science*.

Globe, T. B. (2013). Device makes skipping tv ads easy.

Godes, D. and Mayzlin, D. (2004). Using online conversations to study word-of-mouth communication. *Marketing science*, 23(4):545–560.

Hart (2015). Capturing your consumers in the first 5 seconds.

Harun, A., Razak, M., Nasir, M., and Ali, A. (2013). Freytagâs pyramid: An approach for analyzing the dramatic elements and narrative structure in filem negara malaysiaâs first animated cartoon.

In *IEEE Symposium on Humanities, Science and Engineering Research (SHUSER), Penang, Malaysia.*

Jain, S., Xiong, B., and Grauman, K. (2017). Pixel objectness. *arXiv preprint arXiv:1701.05349.*

Jeong, J., Joung, M. K., and Kim, S. Y. (1998). Quantification of emotion by nonlinear analysis of the chaotic dynamics of electroencephalograms during perception of 1/f music. *Biological Cybernetics*, 78(3):217–225.

Kent, R. J. and Schweidel, D. A. (2011). Introducing the ad ecg. *Journal of Advertising Research*, 51(4):586–593.

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H., Hawk, S. T., and van Knippenberg, A. (2010). Presentation and validation of the radboud faces database. *Cognition and Emotion*, 24(8):1377–1388.

Laurans, G. F. G. (2011). *On the moment-to-moment measurement of emotion during person-product interaction: by means of video-supported retrospective self-report, with some ancillary remarks on other issues in design-related emotion measurement.* TU Delft, Delft University of Technology.

McClure, S. M., Laibson, D. I., Loewenstein, G., and Cohen, J. D. (2004). Separate neural systems value immediate and delayed monetary rewards. *Science*, 306(5695):503–507.

Morwitz, V. G. and Fitzsimons, G. J. (2004). The mere-measurement effect: Why does measuring intentions change actual behavior? *Journal of Consumer Psychology*, 14(1):64–74.

Naik, P. A., Prasad, A., and Sethi, S. P. (2008). Building brand awareness in dynamic oligopoly markets. *Management Science*, 54(1):129–138.

Nodules (2014). *FaceReader.*

O'Donohoe, S. and Young, C. E. (2008). *Emotions, advertising and consumer choice.* Taylor & Francis.

Pease, B. and Pease, A. (2008). *The definitive book of body language.* Bantam.

Petris, G., Petrone, S., and Campagnoli, P. (2009). *Dynamic linear models with R*. Springer Science & Business Media.

Pieters, R., Wedel, M., and Batra, R. (2010). The stopping power of advertising: Measures and effects of visual complexity. *Journal of Marketing*, 74(5):48–60.

Russell, J. A. (1980). A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161.

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

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

Sonnier, G. P., McAlister, L., and Rutz, O. J. (2011). A dynamic model of the effect of online communications on firm sales. *Marketing Science*, 30(4):702–716.

Teixeira, T., Picard, R., and el Kaliouby, R. (2014). Why, when, and how much to entertain consumers in advertisements? a web-based facial tracking field study. *Marketing Science*, 33(6):809–827.

Teixeira, T. S., Wedel, M., and Pieters, R. (2010). Moment-to-moment optimal branding in tv commercials: Preventing avoidance by pulsing. *Marketing Science*, 29(5):783–804.

Texeira, T., Wedel, M., and Pieters, R. (2012). Emotion-induced engagement in internet video ads. *Journal of Marketing Research*, 49(2):144–159.

Viola, P. and Jones, M. J. (2004). Robust real-time face detection. *International journal of computer vision*, 57(2):137–154.

Ward, M. K. and Broniarczyk, S. M. (2011). Itś not me, itś you: How gift giving creates giver identity threat as a function of social closeness. *Journal of Consumer Research*, 38(1):164–181.

West, M. and Harrison, J. (1997). Bayesian forecasting and dynamic models.

WSJ (2015). Youtube: 1 billion viewers, no profit.

Xiao, L., Kim, H.-j., and Ding, M. (2013). An introduction to audio and visual research and applications in marketing. *Review of Marketing Research*, 10:213–253.

Ye, K., Buettner, K., and Kovashka, A. (2018). Story understanding in video advertisements. *arXiv preprint arXiv:1807.11122*.

Appendix

MCMC Sampling Procedure

For convenience, we drop i and j for ease of exposition in some procedures.

1. Generate $y_t^*|Skip_t, \tilde{y}_t, d_t$

$$y_t^*|Skip_t, \tilde{y}_t \sim \begin{cases} N^+(\tilde{y}_t, 1) & if \quad d_t = 0 \,\&\, Skip_t = 0 \\ N^-(\tilde{y}_t, 1) & if \quad d_t = 0 \,\&\, Skip_t = 1 \\ N(\tilde{y}_t, 1) & if \quad d_t = 1 \end{cases}$$

2. Generate $V|e_t^*, \tilde{e}_t$

$$V \sim IW_p(\Psi^V, m^V),$$

$$V|e_t^*, \tilde{e}_t \sim IW_p(\Psi^V + \textstyle\sum_{i=1}^T (e_t^* - \tilde{e}_t)(e_t^* - \tilde{e}_t)', m^V + T),$$

$$\Psi = I_p, m = 10$$

3. Generate $\tilde{e}_t|e_t^*, d_t, \tilde{e_{t-1}}, \Delta, r^e, u^e, W$

We will use forward-filtering and backward-smoothing algorithm for state variables (West and Harrison, 1997). Remember that the observation and system equation are given by

$$e_t^* = \widetilde{e}_t + v_t^e.$$

$$\widetilde{e}_t = \Delta \tilde{e_{t-1}} + r^e d_t + u^e + w_t^e.$$

where $\Delta = \begin{pmatrix} \delta^1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \delta^p \end{pmatrix}$

Let $D_t^e = \{D_{t-1}^e, e_t^*\}$as the information set up to t.

The posterior at t-1 is

$$\widetilde{e}_{t-1}|D_{t-1}^e \sim N(m_{t-1}^e, C_{t-1}^e),$$

where $m_0 = 0_p$, $C_0 = I_p * 10^2$.

The prior at time t is

$$\widetilde{e}_t|D^e_{t-1} \sim N(a^e_t, R^e_t),$$

where $a^e_t = \Delta m^e_{t-1} + r^e d_t + u^e$ and $R^y_t = \Delta C^y_{t-1}\Delta' + W$.

The step-ahead forecast distribution is

$$e^*_t|D^e_{t-1} \sim N(f^e_t, Q^e_t),$$

where $f^e_t = a^e_t$ and $Q^e_t = R^e_t + V$.

Finally, the posterior at time t is

$$\widetilde{e}_t|D^e_t \sim N(m^e_t, C^e_t),$$

where $m^e_t = a^e_t + A^e_t(e^*_t - f^e_t)$, $C^e_t = R^e_t - A^e_t Q^e_t A^e_t$ and $A^e_t = R^e_t(Q^e_t)^{-1}$.

We use backward sampling to obtain draws of $\widetilde{e}_t$ for t=1,2,...T. We first sample $\widetilde{e}_T|D^e_T$ from its conditional distribution. Second, for each t=1,2,...T-1, we sample from

$$\widetilde{e}_t|\widetilde{e}_{t+1,}D^e_t \sim N(h^e_t, B^e_t),$$

where $h^e_t = m^e_t + H^e_t(\widetilde{e}_t - a^e_t)$, $B^e_t = C^e_t - H^e_t R^e_{t+1}, H^e_t$ and $H^e_t = C^e_t\Delta'(R^e_{t+1})^{-1}$.

4. Generate $\widetilde{y}_t|y^*_t, \tilde{y_{t-1}}, \tilde{e_{t-1}}, \beta, \rho, \eta, \phi, u, \sigma^2_w$

Similarly, remember that the observation and system equation are given by

$$y^*_t = \widetilde{y}_t + v_t.$$

$$\widetilde{y}_t = \rho \tilde{y_{t-1}} + \beta'_t \tilde{e_{t-1}} + \eta' z + \phi' Ad + u + w_t.$$

where Ad (Ad_{ij}) denotes the J-dimensional vector of dummy variable where if $Ad_{ij} = 1$, indicates viewer i watched ad j, if $Ad_{ij} = 0$, otherwise.

Let $D^y_t = \{D^y_{t-1}, y^*_t, \tilde{e_{t-1}}\}$as the information set up to t. The posterior at t-1 is

$$\widetilde{y}_{t-1}|D^y_{t-1} \sim N(m^y_{t-1}, C^y_{t-1}),$$

where $m_0 = 0$, $C_0 = 10^2$.

The prior at time t is

$$\widetilde{y}_t | D^y_{t-1} \sim N(a^y_t, R^y_t),$$

where $a^y_t = \rho m^y_{t-1} + \beta' e\tilde{_{t-1}} + \eta' z + \phi' Ad + u$ and $R^y_t = \rho^2 C^y_{t-1} + \sigma^2_w$.

The step-ahead forecast distribution is

$$y^*_t | D^y_{t-1} \sim N(f^y_t, Q^y_t),$$

where $f^y_t = a^y_t$ and $Q^y_t = R^y_t + 1$.

Finally, the posterior at time t is

$$\widetilde{y}_t | D^y_t \sim N(m^y_t, C^y_t),$$

where $m^y_t = a^y_t + A^y_t(y^*_t - f^y_t)$, $C^y_t = R^y_t - A^y_t Q^y_t A^y_t$ and $A^y_t = R^y_t (Q^y_t)^{-1}$.

We use backward sampling to obtain draws of $\widetilde{y}_t$ for t=1,2,...T. We first sample $\widetilde{y}_T | D^y_T$ from its conditional distribution. Second, for each t=1,2,...T-1, we sample from

$$\widetilde{y}_t | \widetilde{y}_{t+1,} D^y_t \sim N(h^y_t, B^y_t),$$

where $h^y_t = m^y_t + H^y_t(\widetilde{y}_t - a^y_t)$, $B^y_t = C^y_t - H^y_t R^y_{t+1}, H^y_t$ and $H^y_t = \rho C^y_t (R^y_{t+1})^{-1}$.

5. Generate $\beta, \rho, \eta, \phi, u | \widetilde{y}_t, y\tilde{_{t-1}}, d_{t,} \sigma^2_w$ (if β_t is time-invariant)

Let $\hat{\beta} = (\beta', \rho, \eta, \phi', u)'$ and $\hat{x_t} = (e\tilde{_{t-1}}', y\tilde{_{t-1}}, z', Ad', 1)'$.Then $\hat{y} = (\tilde{y_1}, ...\tilde{y_T})'$and $X = (\hat{x_1}, ...\hat{x_T})'$

$$\hat{\beta} | \hat{y_t}, \hat{x_t}, \sigma^2_w \sim MVN(\hat{m}, [\Sigma^{-1}_{\hat{\beta}} + \tfrac{1}{\sigma^2_w} \hat{X}' \hat{X}]^{-1}),$$

$$\hat{m} = [\Sigma^{-1}_{\hat{\beta}} + \tfrac{1}{\sigma^2_w} \hat{X}' \hat{X}]^{-1} [\Sigma^{-1}_{\hat{\beta}} \bar{\hat{\beta}} + \tfrac{1}{\sigma^2_w} \hat{X}'(\hat{y} - \hat{X}\hat{\beta})],$$

$$\bar{\hat{\beta}} = 0_3, \Sigma_{\hat{\beta}} = I_3 * 10^2.$$

6. Generate $\sigma^2_w | \widetilde{y}_t, y\tilde{_{t-1}}, d_t, \beta, \rho, \eta, \phi, u$

$$\sigma^2_w \sim IG(a, b),$$

$$\sigma^2_w | \hat{\beta}, \hat{y_t}, \hat{x_t}, \sim IG(a + \tfrac{T}{2}, b + \tfrac{1}{2}(\hat{y} - \hat{X}\hat{\beta})'(\hat{y} - \hat{X}\hat{\beta})),$$

$$a = 1, b = 1.$$

7. Generate $\Delta|\widetilde{e}_t, \tilde{e_{t-1}}, d_t, r^e, u^e, W$

Let $\hat{\theta} = diag(\Delta)$, $\hat{e}_t = (\tilde{e_t} - r^e d_t - u^e)$, $\hat{Z}_t = \begin{pmatrix} \tilde{e_{t-1}}^1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \tilde{e_{t-1}}^p \end{pmatrix}$, $\hat{Z} = \left(\hat{Z}_1{}', ...\hat{Z}_t{}'\right)'$ $\hat{W} = I_N \otimes W$,

$$\hat{\theta}|\hat{e}_t, \hat{Z}_t\hat{W} \sim MVN(\hat{n}, [\Sigma_{\hat{\theta}}^{-1} + \hat{Z}'(\hat{W})^{-1}\hat{Z}]^{-1}),$$

$$\hat{n} = [\Sigma_{\hat{\theta}}^{-1} + \hat{Z}'(\hat{W})^{-1}\hat{Z}]^{-1}[\Sigma_{\hat{\theta}}^{-1}\bar{\hat{\theta}} + \hat{Z}'(\hat{W})^{-1}\hat{e}],$$

$$\bar{\hat{\theta}} = 0_p, \Sigma_{\hat{\theta}} = I_p * 10^2.$$

8. Generate $r^e|\widetilde{e}_t, \tilde{e_{t-1}}, d_t, \Delta, u^e, W$

Similarly,$\hat{D}_t = d_t I_p$

– $r^e|\widetilde{e}_t, \tilde{e_{t-1}}, d_t, \Delta, u^e, W \sim MVN(\hat{n}, [\Sigma_r^{-1} + \hat{D}'(_{\hat{W}})^{-1}\hat{D}]^{-1}),$

$$\hat{n} = [\Sigma_r^{-1} + \hat{D}'(_{\hat{W}})^{-1}\hat{D}]^{-1}[\Sigma_r^{-1}\bar{r} + \hat{D}'(_{\hat{W}})^{-1}(\tilde{e_t} - \Delta\tilde{e_{t-1}} - u^e)'],$$

$$\bar{r} = 0_p, \Sigma_r = I_p * 10^2.$$

9. Generate $u^e|\widetilde{e}_t, \tilde{e_{t-1}}, d_t, \Delta, r, W$

Similarly,$\hat{I}_t = I_p$

$$u^e|\widetilde{e}_t, \tilde{e_{t-1}}, d_t, \Delta, r, W \sim MVN(\hat{k}, [\Sigma_u^{-1} + \hat{I}(_{\hat{W}})^{-1}\hat{I}]^{-1}),$$

$$\hat{k} = [\Sigma_u^{-1} + \hat{I}(_{\hat{W}})^{-1}\hat{I}]^{-1}[\Sigma_u^{-1}\bar{u} + \hat{I}(_{\hat{W}})^{-1}(\tilde{e_t} - \Delta\tilde{e_{t-1}} - r^e d_t)'],$$

$$\bar{u} = 0_p, \Sigma_u = I_p * 10^2.$$

10. Generate $W|\widetilde{e}_t, \tilde{e_{t-1}}, d_t, \Delta, r^e, u^e$

$$W \sim IW_p(\Psi^W, m^W),$$

$$W|\widetilde{e}_t, \tilde{e_{t-1}}, d_t, \Delta, r^e, u_e \sim IW_p(\Psi^W + \textstyle\sum_{i=1}^T(\widetilde{e}_t - \Delta\tilde{e_{t-1}} - r^e d_t - w_t^e)'(\widetilde{e}_t - \Delta\tilde{e_{t-1}} - r^e d_t - w_t^e), m^W + T),$$

$$\Psi^W = I_p, m^W = 10.$$

Simulated Data

In order to test our estimation procedure and demonstrate the identification of the effect on skipping propensity, we conducted a simulated data analysis using the proposed model in Subsection 4.1. The simulation is as follows:

1. We set the number of viewers as $I = 100$, the number of ads they viewed as $J = 3$ and the number of emotions as $P = 2$ respectively. The value of the parameters are set as $\Delta = \begin{pmatrix} 1 & \\ & 1 \end{pmatrix}$, $r^e = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$, $u^e = \begin{pmatrix} 3 \\ 3 \end{pmatrix}$,$V = \begin{pmatrix} 0.25 & \\ & 0.25 \end{pmatrix}$, $W = \begin{pmatrix} 0.25 & \\ & 0.25 \end{pmatrix}$, $\rho = 1$, $\eta = \begin{pmatrix} 0.1 \\ 0.1 \end{pmatrix}$, $u = -2$, $\phi = \begin{pmatrix} 0.5 \\ 0.8 \end{pmatrix}$ $\beta = \begin{pmatrix} 1.3 \\ -3.5 \end{pmatrix}$, $\sigma_w^2 = 0.6$

2. For each viewer i=1,...I, ad $j = 1, ..J$

   - Generate the duration of forced ad exposure $d \sim Multinomial(p)$ where $d \in \{0, 5, 6, 7, ...15\}$ and $p = (\frac{1}{3}, \frac{1}{3}, \frac{1}{30}, \cdots \frac{1}{30})$,

     of d_t are 1, otherwise, are 0.

   - Generate the state vector of emotion $\widetilde{e}_{ijt} \sim MVN(\Delta e_{ijt-1} + r^e d_{ijt} + u^e, W)$, where $\widetilde{e}_{ij0} \sim N(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, W)$,then the observed vector of emotion $e^*_{ijt} \sim MVN(\widetilde{e}_{ijt}, V)$.

   - Generate the state variable of skipping propensity $\widetilde{y}_{ijt} \sim N(\rho y_{i\tilde{j}t-1} + \beta' \tilde{e}_{ijt-1} + \eta' z_{ij} + \phi' Ad_{ij} + u, \sigma_w^2)$, where $\widetilde{y}_{ij0} \sim N(-5, \sigma_w^2)$,then $y^*_{ijt} \sim N(\widetilde{y}_{ijt}, 1)$

   - Then $Skip_{ijt} = I(d_{ijt} = 0)I(y^*_{ijt} > 0)$

We use priors that are similar to those outlined in the estimation procedure in the Appendix. We ran 5,000 iterations in total and discarded the first 2,500 as the burn-in period. The remaining 2,500 iterations were used for the inference. Table 7 reports the true parameters, the estimated posterior means, and the posterior 95% credible intervals. All the posterior means are very close or equal to the true value, and the 95% credible intervals always include the true values. Therefore, all the parameters are successfully recovered.

Parameter	True Value	Posterior Mean	Posterior 95% Credible Interval
V_{11}	0.25	0.25	(0.24,0.26)
V_{12}	0.00	-0.01	(-0.01,0.00)
V_{22}	0.25	0.26	(0.25,0.27)
Δ_{11}	1.00	1.00	(1.00,1.00)
Δ_{22}	1.00	1.00	(1.00,1.00)
r_1^e	2.00	2.02	(1.99,2.05)
r_2^e	2.00	2.03	(2.00,2.06)
u_1^e	-0.30	-0.29	(-0.31,-0.28)
u_2^e	-0.30	-0.29	(-0.30,-0.27)
W_{11}	0.25	0.25	(0.24,0.27)
W_{12}	0.00	0.00	(-0.01,0.01)
W_{22}	0.25	0.23	(0.22,0.25)
β_1	1.30	1.31	(1.30,1.31)
β_2	-3.50	-3.49	(-3.50,-3.47)
ρ	1.00	1.00	(1.00,1.00)
η_1	0.10	0.10	(0.09,0.11)
η_2	0.10	0.09	(0.08,0.10)
ϕ_1	0.50	0.53	(0.47,0.59)
ϕ_2	0.80	0.81	(0.75,0.89)
u	-2.00	-1.92	(-2.07,-1.85)
σ_w^2	0.60	0.59	(0.46,0.88)

Table 7: Parameter Estimates for Simulated Data

To further demonstrate how the model can identify the potential non-transparent skipping decision during forced ad exposure, we plot the latent skipping propensity for one viewing instance based on the simulated data. Figure 10 shows the true skipping propensity trajectory, the estimated skipping trajectory, and the 95% credible intervals during the forced watching period (10s). As we see, all the true values are covered in the 95% credible intervals shade. In particular, the skipping propensity at 2s and 3s, which exceed the skipping threshold, indicates the intended skipping behavior. The estimated skipping propensity can also recover this potential non-transparent behavior as the fitted skipping trajectory across the 0-axis. Now we turn our attention to the performance of the model in the empirical application.

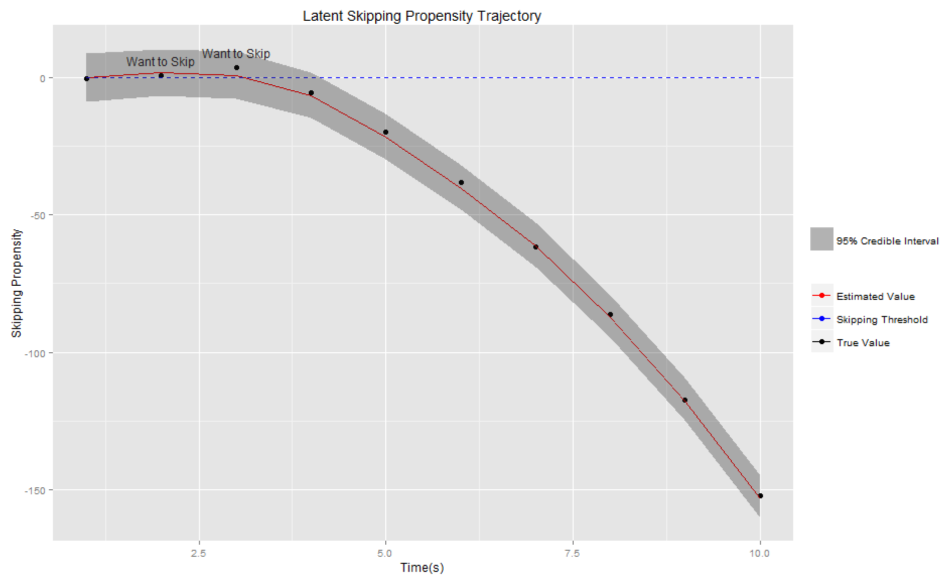


Figure 10: Latent Skipping Propensity Trajectory of y^*_{ijt}

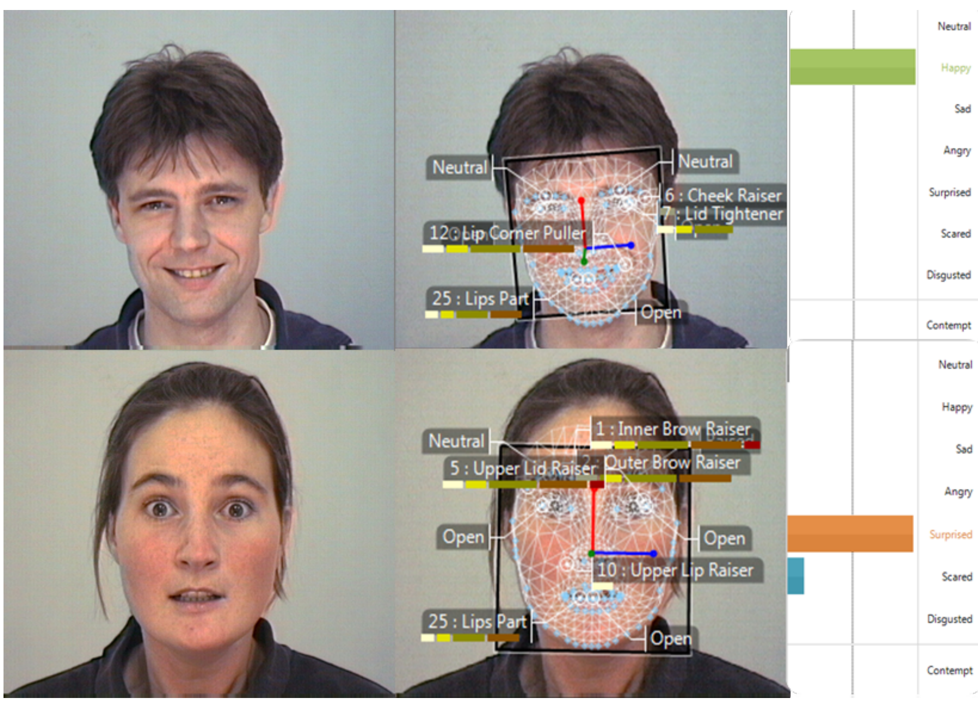


Figure 11: Examples of Facial Recognition