Marketing Science Institute Working Paper Series 2023

Report No. 23-116

The Consumption of Advertising in the Digital Age: Attention & Ad Content

Anthony Dukes and Qihong Liu

The Consumption of Advertising in the Digital Age: Attention & Ad Content*

Anthony Dukes
Marshall School of Business
University of Southern California

Qihong Liu
Department of Economics
University of Oklahoma

January 7, 2023

Abstract

This paper studies the consumption of advertising when attention is costly. Our objective is to understand the advertiser's optimal decision for informational and non-informational content when facing consumers with heavy distractions common in the digital age. We take an equilibrium approach in which an ad is consumed if and only if the content of the ad is worth the viewer's attention. We classify many advertising decisions observed in practice. First, an advertiser can structure content to induce curiosity for continued viewing by showing the information that is relatively *less* likely to resonate with the broad audience. This tactic is evident in clickbait ads and "mystery ads". Second, we find that digital ads, which tend to be skippable, have lower amounts of non-informational content (e.g., entertainment) relative to non-skippable ads. This finding can explain the industry perception that ads on digital media are of lower copy quality than traditional ads.

*The authors are listed alphabetically and contributed equally. They thank Linli Xu, S. Siddarth, participants at the UTD Forms, the ISMS Marketing Science, and the China Marketing International conferences, as well as seminar attendees at City University of Hong Kong, Duke University (Fuqua), New York University (Stern), and the University of Science and Technology of China, for helpful remarks on an earlier draft. Financial support from the Marketing Science Institute (Grant no. 4000790) is gratefully acknowledged.

1. Introduction

Consider the following observations from the advertising practice.

- Clickbait ads have very low CTR's, yet brand information is available only when clicked.
- Ad copy quality is deemed lower for digital ads than for television ads.
- Facebook and Google advise their advertisers to tell the viewer the brand at the start of any video ad.

These observations[1] involve decisions about both informational ad content (brand or product information) and non-informational ad content (e.g., ad copy quality). As we will show in this paper, these and other content strategies can be understood through the economics of attention.

Applying economics to the consumption of advertising is motivated by the observation that consumers in the digital age face increasingly more and tempting distractions away from ads.[2] And because most commercial messages are unsolicited, viewers perceive ads as irritating interruptions interspersed between segments of non-commercial content (e.g., news). Therefore, if any ad is to be consumed, it must provide sufficient value to a viewer to be worthy of her attention. Marketers attempt to provide value through the use of non-informational cues, such as jokes or music, and the careful timing of product and brand information. This leads to the question: *How does an advertiser structure ad content so that it meets communication goals while simultaneously anticipating the consumer's decision of whether to pay attention?*

A volume of prior academic work in marketing explores the factors that affect the consumer's attention to an ad.[3] This large body of research has built a rich understanding of how copy design affects viewers' attention to an ad and its ability to induce recall. Largely overlooked, however, is the viewer's individual incentive to devote attention to the marketer's message. Something as trivial as the consumption of a banner ad comes at the cost of a distraction and the delay in reading an article. And, even if these costs seem small, they are real

[1] Sources: https://contentmarketinginstitute.com/2020/07/clickbait-headlines-good-tactic/, O'Guinn et al. (2009), and internal company research available from the authors.
[2] Boik, Greenstein, and Price (2019) find that, between 2008 and 2013, U.S. households' supply of attention online did not increase despite the proliferation of internet offerings and the increased ownership of viewing devices.
[3] Researchers as far back as Twedt (1952) have studied the impact of various print fonts, colors, and number of words in magazine ads on a reader's ability to recall the seller's message. More recently, using moment-to-moment eye-tracking technology, Teixeira et al. (2010), measured a consumer's attention throughout the duration of video ads.

enough for consumers to disregard portions of many ads. Therefore, for a consumer to devote her attention to an ad message, she must expect some benefit, such as product information or the entertaining aspects of the ad. In this way, the consumption of advertising can be viewed as an equilibrium outcome – one of mutually beneficial decisions.

By ignoring the consumer's incentive to attend to an advertisement, prior work has overlooked two motivations for structuring the ad message, which we identify in this research: the *curiosity inducing* motive and the *short-attention span* motive. The former aims to stimulate consumer's desire to learn more details about the product or advertiser. The latter aims to make the ad effective even when consumers succumb to a distraction before finishing. In addition to identifying these two motivations, this research assesses the conditions under which each one should guide the content decision. A key moderator of interest here is the advertising format: skippable ads versus non-skippable. *Skippable ads*, such as those found on online video-platforms, give viewers an additional source of distraction relative to traditional *non-skippable* ads found on television (Pashkevich et al. 2012). As we show, the commercial medium's use of either ad format can affect the optimal choice of ad content.

Advertising strategies involve two types of informational content: brand impression and product relevance. Brand impression enables the viewer to learn the advertiser's identity and form a favorable or unfavorable impression by showing the brand name and images associated with that brand. Product relevance allows the viewer to determine whether she is in the market for this product category. Brand impression and product relevance are each modeled as a binary signal about its match. Any given ad must contain both signals and both signal realizations must be positive for the viewer to obtain benefits (e.g., purchase the advertised product). It is up to the advertiser to decide in which order the signals are presented to the viewer. Figure 1 (following References) shows how to interpret these two types of signals and the different orders of their presentation. In Panel A the viewer forms a brand impression immediately from the prominent images at the top and subsequently decides whether to spend time assessing the product's relevance by reading the product details underneath. In Panel B, that ordering is reversed. The viewer immediately assesses the product relevance from the displayed information but must take a costly action (i.e., click) to learn about the advertiser.

The clickbait ad in Panel B of Figure 1 can further illustrate the curiosity-inducing motive. It specifies that, to maximize the viewer's curiosity, the advertiser should lead with the signal that is relatively *less likely* to resonate with the viewer. If a viewer experiences a positive "match" from that lower probability signal, here product relevance, then she expects with a higher probability that the other signal will be positive. Hence, she is more incentivized to learn the brand impression and will continue paying attention through the remainder of the ad. That is, only a few viewers will actually click through to learn more, but those that do are precisely those who are highly interested to learn the details about the brand. Interestingly, we also find that the ordering implied by curiosity-inducing motive is always weakly suboptimal for the viewer.

More generally, the curiosity-inducing motive helps unify differences within the marketing literature about whether to lead an ad with the name of an advertiser. Fazio et al. (1992), for example, argue that, in some contexts, it is more effective to make the consumer wait to know the brand. They note the use of "mystery ads" on television that first present details about a product in order to stimulate the consumer's anticipation to learn the brand behind the message, much like in clickbait. By contrast, Baker et al. (2004) offer evidence that an ad is often more effective when the viewer knows upfront the brand being advertised. Our analysis shows how either order can be rationalized in an economic framework and specifies the conditions under which each is more effective. Specifically, a niche brand, which is relatively unlikely to yield a positive match with the average consumer, should lead with the brand information. This induces heightened desire from the small number of brand loyalists to learn more from the ad. And, by contrast, a well-known brand with a new product should do the opposite. By leading with product relevance, those consumers who discover an interest in the new product will be motivated to learn what brand is associated with the desirable product.

The short-attention-span motive, by contrast, recognizes that some advertising goals, such as brand awareness, can be achieved even when most viewers disregard much of the ad. This is illustrated by the magazine ad in Panel A of Figure 1. Here the brand information is prominent at the top of the page allowing the reader to immediately form an impression from the logo, image and catch phrase. Subsequently, she decides to either pursue product details at the bottom of the page or immediately turn the page. Even if most readers skip the details, they still must confront the brand information and leave with an impression. While the advertiser may not generate an immediate sale, a positive impression is additional brand awareness and may result in a future sale or positive word-of-mouth. This motive dominates curiosity-inducing when consumers have low incentives for learning product information and the advertiser finds increasing brand awareness an important goal, as in the case of new products.

The short-attention-span motive is particularly relevant for skippable ads on digital content platforms. This is due to two factors. First, the skippable format, relative to non-skippable, offers viewers a more attractive option to avoid viewing the ad. This means advertisers must compete relatively harder for the viewer's attention in these settings. Second, these platforms often employ an ad pricing system that affords advertisers different rates depending on whether the individual user skipped part of the ad (Dukes, Liu, and Shuai 2022). This *payment-attribution* feature encourages the short-attention-span motive because advertisers can use an ad to generate brand awareness without paying the full exposure cost. In fact, some digital content platforms promote the skippable ad format to advertisers precisely for this benefit.[4]

Finally, we consider the advertiser's use of non-informational content, such as music, a well-known personality, or a funny narrative to combat consumer inattention. These non-informational elements, which we call *entertainment*, represent a direct benefit to a viewer for her attention that is distinct from any informational benefit. Because better entertainment is costly to the advertiser (e.g., hiring a good writer or a celebrity), our modeling objective is to assess the relative return of this investment under different advertising conditions as well as the optimal ordering of the informational signals.[5] Because entertainment combats inattention, it encourages full ad viewing and, therefore, has a higher return with the curiosity-inducing motive. We show further that this effect is more pronounced for the non-skippable format than for the skippable format. Therefore, all else equal, a content strategy for a skippable ad will have lower equilibrium investment in entertainment relative to a non-skippable ad. This finding is consistent with the notion that high-price ad agencies (e.g., Madison Ave.) are still associated with TV advertising and explains the industry perception that ads on traditional print and broadcast media

---

[4] See Google's promotion of YouTube's *TrueView* ads https://support.google.com/youtube/answer/2375464?hl=en.
[5] We do not attempt to capture emotional or logical cues explicitly. Rather, we embody any generic direct benefit from viewing the ad itself. For insights on explicit ad copy decisions, see Chandy et al. (2001).

are generally of better copy quality relative to digital ads (O'Guinn et al. 2009, p225), many of which are skippable.

Advertising is one of the most studied areas in the marketing literature. Much of the work tries to understand what executional features make an ad effective (e.g., Twedt 1952, Diamond 1968, Stewart and Koslow 1989, and Tellis et al. 2000, Bruce et al. 2017). We complement this large body of work by studying the viewer's individual incentives to pay attention to ad content and then exploring how these incentives vary across different media formats. We do not seek to offer a theoretical explanation of how advertising works (Vakratsas and Ambler 1999). Rather, we draw insights for marketers to improve their communications with consumers facing so many distractions in the digital age. In this way, our work relates to three important sub-streams within the advertising literature: *viewer attention*, *ad content*, and *the role of the medium*.

The marketing literature on advertising solidly recognizes the necessity of *viewer attention* in ad effectiveness (Barry and Howard 1990) and its moderating factors (Teixeira et al. 2010). It examines, for example, how elements of ad copy affect attention as measured by brand liking and recall (Singh et al. 1988, Walker & Dubitsky 1994, and Du Plessis 1994), eye movements (Wedel & Pieters 2008 and Teixeira et al. 2010) and other neurophysiological reactions (Venkatraman et al. 2015). We complement this body of work by incorporating the consumer's economic motivation for attending to the ad as the manager considers market and media variables when making decisions about the placement of brand and product information to achieve communication goals. Another literature has recognized the role of limited attention in marketing communication in equilibrium contexts (van Zandt 2004).[6] Some of that literature focuses on how attention decisions affect pricing outcomes (Anderson & de Palma 2012 and Zhu & Dukes 2017) whereas our focus is on advertising content and consumption. A closely related work to ours is Jerath & Ren (2021), which studies the consumer's allocation of attention across favorable and unfavorable product information *within* an ad or set of product reviews. Our focus, by contrast, is on the consumer's optimal allocation of attention between ad viewing and some other activity, such as watching video content or checking email.

---

[6] A separate literature from economics on *rational inattention* (e.g., Sims 2003, Reis 2006 and Matejka 2016) explains why agents rationally ignore information.

Related to the decision of attending to an ad is the notion of ad avoidance, which comprises ad avoidance technologies (AAT), channel zapping, and skippable ads.[7] Recent work on AAT, such as digital video recorders and ad blockers, examine the decision to avoid all ads and the implication of this technology on outcomes in the advertising market (Anderson & Gans 2011, Ghosh & Stock 2012, Johnson 2013, Chen and Liu 2021, and Gritckevich et al.). Our work, by contrast, starts at the individual level of the ad-viewer interaction, particularly when the viewer's decision to attend to an ad is conditioned on the initial information provided at the start of the ad. This relates to the work on ad avoidance via "zapping", in which the viewer avoids a television commercial by switching to another channel. That research empirically identifies factors that induce switching (Siddarth & Chattopadhyay 1998) and its timing (Gustafson & Siddarth 2007). Similarly, we are interested in factors that affect a consumer's decision to continue attending to an individual ad. However, in this paper we take an equilibrium approach to understand the economic mechanisms behind these factors. Within the realm of skippable ads, Pashkevich et al. (2012), Belanche et al. (2017), and Hao et al. (2020) explore the drivers of engagement and satisfaction in skippable ads. Pashkevich et al. (2012), in particular, find that (i) skippable ads are preferred by viewers over non-skippable; and (ii) those who do not skip are more engaged and attentive to the ad. Both findings are consistent with and explained by our model. Finally, Dukes et al. (2022) study the ad-viewer interaction from the perspective of a two-sided market in order to assess how a switch to the skippable format affects ad volume and platform revenue. As such, it disregards the strategic incentives of advertisers to structure their content as a function of the ad format, as we do here.

Another important sub-stream in the advertising literature examines the selection of *ad content*. Most relatedly is McGranaghan et al. (2021), which empirically substantiates the premise of our work by identifying how ad content choices can affect viewer attention and its subsequent effectiveness. An older stream of work on ad content was particularly concerned with the portion of the ad that was actual product information (Resnik and Stern 1977) and how this portion varies across time and media (Stern et al. 1981, Abernathy & Frank 1996). Much of that work suggests that non-informational cues are persistent in advertising and that little content is

[7] The consumer's attention decision also plays an important role in advertisers' use of native advertising. Specifically, advertisers assume that consumers are more inclined to attend to the advertiser's message if it appears "native" to the platform's organic content (Sahni & Nair 2020 and Wu, Gal-Or, & Geylani 2022).

related to price with detrimental effects on consumer choice and competitive outcomes (Abernathy & Frank 1996). Follow-up work in economics tries to explain this persistent observation as it relates to equilibrium pricing through money-burning (Nelson 1974, Kihlstrom & Riordan 1984, and Milgrom & Roberts 1986), consumer search (Anderson & Renault 2006 and Mayzlin & Shin 2011) or consumer segmentation (Soberman & Xiang 2021). In contrast to that prior work, we abstract away from product pricing to focus on how informational and non-informational cues affect the consumer's decision to pay attention. There is also a stream of work that examines ad content within the duration of an ad, which shows that the frequency of brand cues significantly affects viewer reactions. While the presence of brand cues is necessary, too frequently presenting them can divert the viewer's attention (Teixeira et al. 2010) and even compromise sales (Bruce et al. 2020). We build on this stream of work by showing how the order of product and brand information affects the viewer's incentive to learn more about the advertiser. In other words, we examine how the placement of brand information helps generate a viewer's curiosity about an ad's content.

Finally, the advertising literature studies the *role of the medium* through which an ad is delivered (e.g., Dukes and Gal-Or 2003, Wilbur 2007, Godes et al. 2009). A lot of this work is focused on helping managers in media planning and targeting (Lin et al. 2013), multimedia ad campaigns (Danaher and Dagger 2013), and synergies across media (Naik and Raman 2003). This work predominantly treated ad content as fixed, whereas the current work is aimed at understanding how ad content and consumer attention interact with the commercial medium.[8] For our purposes, we are interested in how different media properties moderate the optimal content strategy by changing the incentives for attention. This moderator is particularly relevant for media that provide the option to skip ads and provide an additional incentive for the viewer to divert her attention (Belanche et al. 2017 and Dukes et al. 2022).

2. Model

The model consists of two types of agents: viewers and one advertiser. A viewer, sometimes also referred to as a consumer, interacts with an advertisement over two periods, $t = 1,2$. Consumers differ only in regard to their attention cost, $c > 0$, which the consumer incurs when consuming

[8] A recent exception is Lin (2022), which examines the consumer's attention to ads from the perspective of the advertising medium and its choice of mixing commercial and non-commercial content.

the remainder of the ad in period $t = 2$. The fact that ads typically interrupt a viewer's desired platform content indicates that an early portion of the ad, even if short, may be "forced" upon the viewer. At this point, in period 1, the viewer must decide how to allocate her attention to the remainder of the ad, which occurs in period 2. This approach limits our ability to say anything about what makes a viewer initiate watching an ad and, correspondingly, how the advertiser's decisions affect the ad's reach. Therefore, we assume that the ad has a normalized reach of one.[9]

An advertiser provides exactly two types of informative content and a level of non-informative content. Informative content comes in the form of two binary signals, each representing a different dimension of information for the consumer that is potentially valuable for the consumer. We interpret one as the *brand impression* and the other the *relevance indicator*. In period $t = 0$, the advertiser decides in which order the signals are presented to the viewer. Simultaneously, the advertiser invests in a level of non-informative content, $e > 0$, at a cost ke^2. The value e, which we call *entertainment*, is a direct benefit to the viewer for watching.[10] This benefit is independent of informational outcomes. The value of e is assumed fixed until Section 4.

The *brand impression* (I) can be viewed as identifying the advertiser and giving the viewer valency, without giving immediate benefit to the viewer. That is, upon seeing this signal, the viewer learns the identity of the advertiser and then forms an impression, which is favorable ($I = 1$) with probability $\rho_I \in [0,1]$. Otherwise, the viewer has an unfavorable impression of the advertiser ($I = 0$). The *relevance indicator* (R), by contrast, gives product information that the viewer can use to assess whether the advertised product is relevant for her present needs. The viewer finds the advertiser relevant ($R = 1$) with probability $\rho_R \in [0,1]$. Otherwise, the advertiser is not relevant ($R = 0$). Figure 1 illustrates the interpretations of signals I and R in an example from a magazine ad. If the brand impression and relevance indicator signals are jointly successful ($I = 1$ and $R = 1$, in either order), then the viewer obtains a direct benefit $u_s > 0$, stemming from a purchase or some beneficial action. Thus, upon viewing both signals, the viewer receives benefit with probability $\rho_I \rho_R$. Otherwise, the viewer receives little to no benefit

[9] We discuss broader implications of this assumption at the end of Section 5.
[10] We use the nomenclature "entertainment" simply to give context to non-informative content. However, it can be interpreted more broadly to also include serious or disturbing ad content – anything that compels the viewer to watch for non-product benefits.

from the advertiser. She may still receive in period $t = 2$ a benefit from the advertisement itself, $e > 0$, in the form of entertainment, as chosen by the advertiser in period $t = 0$.

We focus on what we believe are two natural dimensions of information (brand and product), each represented by a match value (I or R). An important condition is that, for any ad, the viewer is unaware or has forgotten whether the advertised product is relevant and must be told whose product it is. We acknowledge that this is an abstract representation of informational ad content, which cannot directly be mapped to content as interpreted by a practitioner. By the same token, this abstraction affords benefits. When combined with the specification of outcome payoffs, this representation allows a sufficiently rich interpretation that can be applicable to many advertising situations.[11] As we will demonstrate subsequently, the payoff structure (for weak versus strong conversions) is used to augment our interpretation of these match values as ad content (brand impression and product relevance). This indirect approach to interpreting content endows the model's agents (viewers and the advertiser) with the strategic tradeoffs we seek to study (attention and content) with the analytical flexibility to derive practical insights.

The advertiser benefits from ad consumption through an event we call a *conversion*. While the benefit of a conversion has broad interpretations, we avoid assessing advertiser strategy using the classic advertising metrics such as reach, broadness of appeal, and other campaign related measurements. There are two types of a conversion, each of which constitutes a benefit to the advertiser stemming from an action from the representative viewer. A *weak conversion* gives the advertiser $V_w \geq 0$ and requires only that the consumer has a positive impression of the advertiser ($I = 1$). The viewer receives a reduced benefit, $u_W < u_S$, which we normalize to 0.[12] A *strong conversion* gives the advertiser $V_S > V_w$ and requires that *both* signals are successful – a favorable impression that is also relevant (with probability $\rho_I \rho_R$). See Table 1 (following References) for a summary of the configurations of signals, conversions, and payoffs. By construction, the advertiser's payoff from conversions are differentially aligned with the consumer's.[13] This serves two purposes in our model. First, it generates strategic interactions

[11] This formulation permits a wide variety of situations including ads from a collective of brands (e.g., trade association) or a non-profit (e.g., seeking donations).
[12] While this normalization does not undermine the intuition or economic forces behind our results, additional insights arise in the case of $u_W > 0$. See footnotes 14, 18 and 21 for more discussion.
[13] We assume that the viewer has a smaller weak-to-strong benefit ratio relative to the advertiser: $u_W/u_S \leq V_W/V_S$. The simplest case of this condition is $u_W = 0$. Further, it maintains the desired property of contrasting incentive

between the advertiser in his choice of signal order and the viewer's choice to pay attention to the second part of the advertisement. Second, this payoff structure has flexible interpretations relating to different goals of advertising. For example, a strong conversion can be interpreted as a purchase or some other consumer action that mutually benefits the consumer and advertiser. Alternatively, one can interpret a weak conversion as the advertiser achieving an increased brand awareness or building customer-based brand equity. In such cases, the advertiser receives some benefit even if the viewer obtains relatively little.[14]

Under our interpretation of ad content, we use the exogenous probabilities ρ_I and ρ_R to represent market factors reflective of existing marketing situations, which the advertiser takes as given. For instance, a small (large) ρ_I, can represent a niche (mass appeal) brand. A small (large) ρ_R can be interpreted as a small (large) market need. This interpretation is to help guide our ideas, though we acknowledge it does not capture all the complexity of conditions in the advertising practice. With such market interpretations, we assume both viewer and advertiser view these probabilities as common knowledge. In particular, the agents know the ordering of these two market parameters ($\rho_R < \rho_I$ versus $\rho_R > \rho_I$), which plays an important role in the interpretation of our results. We refer to the signal with the smaller value of ρ as the "weaker signal" as it has a lower message match probability.

The advertiser decides in what order to present the signals – either I or R first, followed by the other (R or I, respectively). While we present the signal order as a sequence in time, the model loosely permits an interpretation of simultaneous presentation, say a print magazine ad, in which one signal is made explicitly prominent so that the viewer attends to it first. Figure 1 illustrates this interpretation of signal order, with I first in Panel A and R first in Panel B. The strategic aspect of the order decision is its effect on the consumer's incentives to pay attention to the entire ad. Signals are revealed with certainty in the period they are presented and, therefore,

---

alignments among strong and weak conversions. Specifically, viewer and advertiser incentives are aligned in a strong conversion, but not for a weak conversion, which benefits only the advertiser.

[14] To further help with interpreting the advertiser's payoffs from a weak conversion, consider a viewer who is not in the market for a particular product ($R = 0$) but gains a positive impression of the advertising brand ($I = 1$). This viewer then tells a friend about it through word-of-mouth and the friend later buys the advertised product. Here, the advertiser obtains a lagged benefit ($V_W < V_S$) and the viewer obtains nothing ($u_W = 0$), by assumption. However, the case of $u_W > 0$ has a meaningful interpretation with interesting comparative statics. For example, the consumer with a weak conversion ($I = 1$, but $R = 0$) may derive benefit from making a recommendation online or to a friend (Campbell et al. 2016). See footnotes 18 and 21 for implications.

the advertiser has no incentive to exclude either signal in an ad. Based on the realization of the first signal (favorable or unfavorable), the consumer decides whether to attend to the second one. Receiving the second signal costs the consumer $c > 0$ and reflects her effort in paying attention to and concentrating on an ad.[15] Formally, we normalize the attention cost in the first period to zero so that all viewers consume the first signal in period $t = 1$ and have positive attention costs for attending to the second signal in period $t = 2$. These costs are distributed uniformly: $c \sim U[0, \bar{c}]$. The upper end of the support, $\bar{c} > 0$ represents the consumer whose attention cost is so high that she never engages with the second part of the ad. Watching television at a sports bar with friends may have a higher value of $\bar{c}$ than television viewing alone at home. The elements of attention cost and of signal order described so far can be illustrated in the magazine ad shown in Figure 1. In that example, the brand impression (I) is conveyed first because it is prominently displayed at the top of the ad and accessible with little viewer effort. The lower portion of the ad contains detailed product information, which conveys relevance (R), if the viewer spends her time (attention cost) reading.[16]

All viewers anticipate non-commercial content after the advertisement, which gives a utility of $w > 0$, if viewed in period $t = 2$. If the viewer watches the second part of the ad, and does not consume content until after period 2, her utility from non-commercial content is discounted by a factor of $\delta > 0$. By watching the second part of the ad, viewers receive direct utility $e \geq 0$ from the ad in the form of entertainment. This can be a joke, a story from a good-looking celebrity, or compelling music. Such elements are set up for the viewer in period $t = 1$, but not fully enjoyed until the second part of the ad. In Section 4 we consider the advertiser's choice of e, but for now it is considered fixed and common knowledge.

Finally, we want to understand how the ad format moderates the optimal content. We consider two ad formats. The *non-skippable* ad format (N) defines that the viewer must wait until after period 2 to consume content, which delivers a discounted utility of δw. By not attending to the second message (period $t = 2$), the viewer saves her attention cost c and receives a utility of zero from the ad. Alternatively, in the *skippable* ad format (S), the viewer can skip the second

[15] We assume here that attention costs do not depend on signal realizations but relax this restriction in Section 6.2.
[16] The normalization of the first period attention costs overlooks the possibility that viewers incur costs at the start of the ad simply due to the disruption from enjoying the platform's content. These costs have practical consequences for the initial size of the ad's audience.

period's signal and enjoy the undiscounted utility of platform content. To be explicit with nomenclature, a viewer can "ignore" the second part of the ad in the non-skippable format but can actually "skip" this part and directly enjoy undiscounted platform content with the skippable format.

Before moving to the formal results, it is helpful to draw a connection between our equilibrium conceptualization of attention to that of prior conceptualizations. Specifically, Venkataraman et al. (2015) make the distinction between "top-down" and "bottom-up" notions of attention. The former is captured by a viewer's decision to continue and obtain the second signal. The latter is captured by revelations in the first period (either a favorable informational signal or the expectation of entertainment) that raise the benefit of the second part. This form of state dependence within the consumption of the ad is also noted in Teixeira et al. (2010), who show how the early appearance of brand information (e.g., logo and brand name) affect the consumer's subsequent level of attention. Similarly, in our framework, it is the consumer's incentive to either learn more about the advertiser or to be entertained that determines whether she stays tuned in.

3. Basic Motives for Informational Content

The two basic motives for informational content can be demonstrated with a simplified setting: fixed entertainment under the non-skippable format. We first describe the viewer's attention decision under the two signal orderings (R versus I) and then illustrate the curiosity-inducing and short-attention span motives as they relate to the advertiser's profitability under each. Finally, we demonstrate how these motives are affected by the skippable ad format. In Section 4, we show how the inclusion of endogenous entertainment affects content choices. Table 2 (following References) collects and describes the model's parameters.

3.1 Viewer's Attention Decision

Suppose the advertiser leads with I.[17] Because the first signal is costless for the viewer to acquire, all viewers are exposed to it. If the viewer pays attention in the second period, then she can expect a utility of

[17] To aid this discussion, the consumer's decision trees are provided in Figures A2 and A3 of the Appendix.

$$U_I^N(c) = \begin{cases} \rho_R u_s + (e - c) & \text{if } I = 1 \\ e - c & \text{otherwise,} \end{cases} \quad (1)$$

for a fixed level of entertainment $e \geq 0$. If the viewer does not pay attention in the second part of the ad, her utility is normalized to zero. Thus, we can define an attention cost threshold $\hat{c}_I^N$ such that any viewer with $c < \hat{c}_I^N$ attends to the second signal, which is favorable (i.e., $R = 1$) with probability ρ_R. Specifically, setting (1) to zero gives two thresholds, $\hat{c}_I^N = \rho_R u_s + e$, if the first signal was favorable, and $\tilde{c}_I^N = e$, otherwise. Note that if $I = 0$, then no conversion is possible and the corresponding threshold ($\tilde{c}_I^N$) is irrelevant to the advertiser and hence is ignored. The threshold $\hat{c}_I^N$ increases in e, but the viewer receiving $I = 1$ is further motivated to pay attention from the value of product relevance information, which is the expected value of a strong conversion, $\rho_R u_s$.

If the advertiser leads with R in the non-skippable format, her utility is

$$U_R^N(c) = \begin{cases} \rho_I u_s + (e - c) & \text{if } R = 1 \\ e - c & \text{otherwise,} \end{cases} \quad (2)$$

which defines thresholds $\hat{c}_R^N = \rho_I u_s + e$, if the first signal is favorable (i.e., $R = 1$), and $\tilde{c}_R^N = e$ otherwise.[18] If $R = 0$, then the advertiser still benefits from a weak conversion if the viewer attends to the second part of the ad and leads to a favorable brand impression ($I = 1$).

We summarize the viewer's attention decision using cutoffs $\hat{c}_R^N$, $\hat{c}_I^N$ and $\tilde{c}_R^N$, the exact expressions of which are in Lemma A1 of the Appendix.[19] The benefit of continued viewing when the first signal is successful ($\mathcal{s} = 1$), as defined by the attention threshold $\hat{c}$, is increasing in $\rho_{I/R} u_S$ and e, the expected informational benefit of attention ad and the non-informational benefit, respectively. These benefits are traded-off against attention cost $c > 0$. Whenever the first signal is $\mathcal{s} = 0$, there can be no informational incentive. Only entertainment from the ad can inspire attention. In fact, the advertiser obtains no benefit from consumer attention when the first

[18] Suppose $u_W > 0$. It can be shown that, if the first signal is either $R = 1$ or $I = 0$, then the consumer's attention decisions are identical to the case of $u_W = 0$. If $R = 0$, then the attention cost threshold is $\tilde{c}_R^N = (1 - \rho_I) u_W + e$. Thus, relative to the case of $u_W = 0$, consumers are more inclined to pay attention under a content strategy of R because of the possible benefit from a weak conversion. If $I = 1$, then the attention cost threshold becomes $\hat{c}_I = \rho_R(u_S - u_W) + e$; consumers are less inclined to pay attention under a content strategy of I because of the lower incremental benefit of attention.

[19] Under our model conditions, all cutoffs are less than $\bar{c}$, so that a positive portion of viewers do not attend to the second part of an ad. Formal details are found in Lemma A1 of the Appendix.

signal is 0, except in the non-skippable format with $R = 0$. In that case, $\tilde{c}_R^N/\bar{c}$ portion of consumers stick around purely for entertainment giving the advertiser a weak conversion with a probability of ρ_I.

3.2 Advertiser Strategies with the Non-Skippable Format

The advertiser decides how to structure informational ad content by choosing the order of signals, leading either with R or with I. In arriving at the advertiser's objective function, we assume that the *advertisement's reach* is constant and normalized to one.[20] Let $r \geq 0$ be the price the advertiser pays for an ad. Under the non-skippable format, the advertiser pays r regardless of whether the viewer attends to the entire ad. The profit for the advertiser under the non-skippable format, as a function of content:

$$\pi_I^N = \rho_I\{\hat{c}_I^N[\rho_R \cdot V_S + (1-\rho_R) \cdot V_W] + (\bar{c} - \hat{c}_I^N) \cdot V_W\}/\bar{c} - r - ke^2; \tag{3}$$

$$\pi_R^N = [\rho_R \rho_I \tilde{c}_R^N \cdot V_S + (1-\rho_R)\rho_I \tilde{c}_R^N \cdot V_W]/\bar{c} - r - ke^2, \tag{4}$$

where ke^2 (with $k > 0$) is the advertiser's investment cost for entertainment level e, which takes on relevance in Section 4.[21]

To analyze the optimal informational content, define $\Delta^N \equiv \pi_I^N - \pi_R^N,$ as the incremental profit from leading with I.[22] Then the condition for leading with I is given as follows:

$$\Delta^N > 0 \Leftrightarrow \rho_R u_s V_S(\rho_R - \rho_I) + (\bar{c} - \rho_R^2 u_S - e)V_W > 0. \tag{5}$$

The above condition reveals the two distinct drivers of the advertiser's optimal content strategy. The first term in (5) indicates the incremental profit from a strong conversion when going from R to I. The second term defines the incremental profit from a weak conversion. We study each of these two drivers in isolation. We first shut down the second driver by setting $V_W = 0$, which implies that all advertiser profits come from a strong conversion and requires the

[20] The implications of this assumption are discussed at the end of Section 5.
[21] While the assumption of zero consumer benefit for a weak conversion ($u_W = 0$) simplifies the subsequent analysis, it is useful to recognize how $u_W > 0$ affects the advertiser's preference for signal order. Due to its effect on $\tilde{c}_R^N$ and $\hat{c}_I^N$ (footnote 18), any increase in u_W gives the advertiser marginally more incentive to lead with R over I. We thank an astute reviewer for recognizing this model feature.
[22] Alternatively, we can allow the advertiser to strategically "mix" signals. (Fully processed both signals in one period runs counter to our basic premise that attention is scarce.) However, as (5) makes clear, mixing is weakly optimal and only when $\pi_I^N = \pi_R^N$.

viewer to watch the entire ad. Then we study the second driver by shutting down the first one by setting $\rho_R = \rho_I$ and allow $V_W > 0$.

First suppose $V_W = 0$. Under this condition, the advertiser obtains a benefit from an ad only from a strong conversion. Noting that a strong conversion also gives benefit to the viewer, the advertiser's incentive is aligned with the viewer's. As such, the advertiser can induce continued viewing by maximizing the viewer's expected benefit from conversion. This is seen directly from the condition (5) when $V_W = 0$ from the fact that $\Delta^N > 0$ if and only if $\rho_R > \rho_I$. The advertiser leads with the weaker signal so that the viewer who has had a successful signal in the first period, has a larger expected benefit for attending to the second part of the ad. This motive capitalizes on the viewer's incentive to acquire more information in the second signal. Therefore, we refer to the advertiser's motive to exploit the viewer's desire for more information as the *curiosity-inducing motive*. This motive relies on the viewer's desire for a strong conversion and applies uniquely to that advertiser goal. It is activated by ordering information (brand impression or generic relevance) with the weakest signal first: $\mathrm{argmin}_{\mathscr{s}}\{\rho_{\mathscr{s}}\}$.[23]

To examine the second motive of the advertiser's informational content decision, suppose $V_W > 0$. To illuminate this in isolation, temporarily impose the condition $\rho_R = \rho_I$. It then immediately follows from (5) that $\Delta^N > 0$. The advertiser unambiguously prefers to lead with I because it increases the possibility of a weak conversion. Leading with I means that a weak conversion can occur regardless of how much attention the viewer pays to the remainder of the ad. As such, we refer to this as the *short-attention-span motive* for structuring information ad content. The short-attention span motive relies on the consumer's tendency to not attend to the full ad and thereby delivers a weak conversion in the first part of the ad in which there is the maximal audience.

Now we consider the optimal content strategy more generally, off the 45°-line ($\rho_R = \rho_I$) and when $V_W > 0$. Define the indifference curve $\Gamma^N(\rho_I, V_W) \equiv \{\rho_R \in [0,1] | \Delta^N = 0\}$, which is the value of ρ_R that equates $\pi_I^N = \pi_R^N$ for any $\rho_I \in [0,1]$ and $V_W \geq 0$. It is easily shown using (3)-(4) that $\Gamma^f(\rho_I, V_W)$ defines a curve in $[0,1]^2$ that is represented by an increasing function of

[23] It is noteworthy that the advertiser's optimal order on informational content may not align with the viewer's preferences. Separate analysis shows that the viewer always weakly prefers leading with the *stronger* signal because it is more diagnostic for her attention decision. This has important implications for content platforms, which we discuss in Section 6.2.

ρ_I. Furthermore, we know from the above discussion that for any ρ_I, $\Gamma^N(\rho_I, 0) = \rho_I$ and $\Gamma^N(\rho_I, V_W > 0) < \rho_I$. The equilibrium for the case of fixed $e > 0$ is fully characterized formally in the result below and graphically depicted in Figure 2 (following References).

> **Proposition 1** (Optimal Signal Order with Exogenous e) *Suppose $e < \hat{w}$, e fixed, with the non-skippable ad format (N). Then, under the conditions of Proposition A1 (in Appendix), $\Gamma^N(\rho_I, V_W)$ has the following equilibrium properties.*
>
> (i) $\pi_I^N > (<) \pi_R^N$, *whenever* $\rho_R > (<) \Gamma^N(\rho_I, V_W)$;
>
> (ii) $\Gamma^N(\rho_I, V_W)$ *is the 45°-line for* $V_W = 0$ *and is decreasing in* $\frac{V_W}{u_S V_S} > 0$.
>
> (iii) *For $V_W > 0$, the short-attention-span motive (lead with I) is dominant for the advertiser's optimal content strategy whenever $\rho_R \in \left(\Gamma^N(\rho_I, V_W), \Gamma^N(\rho_I, 0)\right)$. Otherwise, the curiosity-inducing motive is dominant (lead with* $\mathrm{argmin}_s\{\rho_s\}$*).*

Part (i) confirms the interpretation of Γ^N as indifference curves in (ρ_I, ρ_R)-space.[24] Understanding part (ii) relies on the interpretation of the term $V_W/(u_S V_S)$. This expression measures the value of the short-attention-span motive relative to the curiosity-inducing motive. The larger it is (e.g., as V_W increases), the greater the advertiser's gain from a weak conversion by leading with I. Conversely, larger values of $u_S V_S$, the stronger is the incentives of both viewer and advertiser to shoot for a strong conversion by invoking the curiosity motive and leading with R. This intuition summarizes part (iii) and helps explain the impact of a change to the skippable ad format.

3.3 Advertiser Strategies with the Skippable Format

Under the skippable format, the viewer's attention decision is identical to the non-skippable format except in one aspect: the cost of pursuing a second signal. For skippable ads, the viewer acquires the second signal at a cost $c + \hat{w}$, where $\hat{w} \equiv (1 - \delta)w$ is the cost in delaying the desired content. Therefore, we derive the attention cost cutoffs under the skippable format simply by accounting for the additional option to obtain platform content sooner: $\hat{c}_I^S = \rho_R u_S + e - \hat{w}$ and $\hat{c}_R^S = \rho_I u_S + e - \hat{w}$. This distinction implies that when $R = 0$, all viewers skip to

---

[24] Proposition A1 gives general parametric conditions to ensure that we can rely on the heretofore established mechanisms when presenting results. Doing so, however, implies caution with claims along the axes in Figure 1.

content under an assumption $\hat{w} > e$. This assumption is maintained throughout the analysis and guarantees that no irrelevant advertisement is preferable to the intended content on the medium.

On the advertiser-side, the skippable format may include the payment attribution property, which means that the advertiser pays $r \geq 0$ only when the viewer attends to the entire ad.[25] The advertiser's profits, for each content order (I or R) under the skippable ad format derive similarly as before, except that the advertiser does not pay when the consumer skips. This leads to the following expressions:

$$\pi_I^S = \rho_I\{\hat{c}_I^S[\rho_R \cdot V_S + (1-\rho_R) \cdot V_W] + (\bar{c} - \hat{c}_I^S) \cdot V_W\}/\bar{c} - r\,\rho_I(\hat{c}_I^S/\bar{c}) - ke^2; \tag{6}$$

$$\pi_R^S = \rho_R(\rho_I V_S - r)\hat{c}_R^S/\bar{c} - ke^2. \tag{7}$$

To understand how the ad format moderates the advertiser's content strategy, consider condition for the optimality of leading with I

$$\Delta^S > 0 \Leftrightarrow (\rho_R\rho_I u_S V_S - r\hat{w})(\rho_R - \rho_I) + \rho_I[\bar{c} - \rho_R(\rho_R u_S - \hat{w} + e)]V_W > 0, \tag{8}$$

which is analogous to condition (5) of the non-skippable format. As before, $V_W = 0$ is necessary to shut down the short-attention-span motive. However, unlike the earlier case, it is required that $\rho_R\rho_I V_S > r(\hat{w}/u_S)$.[26] Without this latter condition the advertiser will lead with the stronger signal in order to reduce ad costs by discouraging the viewer from paying attention, which is an uninteresting scenario. Therefore, we will assume henceforth that it holds. As before, when $V_W = 0$, the optimal content strategy is driven exclusively by curiosity-inducing motive: lead with the signal with lowest match probability.

And as is the case with the non-skippable format, $V_W > 0$ is necessary for the existence of the short-attention-span motive, which is stronger under the skippable format. To see this, consider the relative attractiveness of I over R, as measured by $\Delta^S - \Delta^N$, is stronger for skippable

---

[25] To keep the analysis focused on the viewer-advertiser interaction, we do not model decisions in the advertising market, such as pricing by the platform. Therefore, we maintain r as exogenous and fixed across formats. Even though this unit rate is likely to differ across formats in practical ad markets, our assumption of fixed r is innocuous as long as we avoid comparing advertiser profits across formats.

[26] The left-hand side of this inequality reflects the advertiser's expected benefit from the viewer's pursuit of the second signal and the right-hand side the expected savings for the advertiser when the viewer skips the second part of the ad. This savings is proportional to the price of a fully viewed ad, $r > 0$, and the viewer's trade-off from pursuing the ad, $\hat{w}/u_S$. Thus, the condition means that the payment attribution incentive of the advertiser is inferior to that of a strong conversion.

ads. According to (5) and (8), $\Delta^S - \Delta^N > 0$ if (i) $V_W > 0$ and (ii) $r(\rho_I - \rho_R) > 0$. In the former condition, the advertiser cares about a weak conversion, as per the short-attention span motive. The second condition furthers the advertiser's incentive to lead with I to capitalize on the payment attribution property of skippable ads. When the viewer skips the ad, the advertiser saves ad costs if $r > 0$. Leading with I when $\rho_R < \rho_I$, encourages this cost savings because the viewer is less pursuant of the second signal, R. Proposition 2, part (i), formalizes this result.

Proposition 2 *Let $\Gamma^S(\rho_I, V_W)$ be the advertiser's indifference curve under the skippable ad format, (defined analogously to Γ^N). Then, under the conditions of Proposition 1 and for any* $r \geq 0$,

(i) *Leading with I is more attractive under the skippable format:* $\Gamma^S(\rho_I, V_W) < \Gamma^N(\rho_I, V_W) < \rho_I$;

(ii) *Raising ad costs makes leading with I more attractive only for the skippable format:* $\partial\Gamma^N/\partial r = 0 > \partial\Gamma^S/\partial r$.

Figure 2 depicts the equilibrium strategy for informational content under each format as described in Propositions 1 and 2. Further, as expressed by part (i) of Proposition 2, there is a region in the (ρ_I, ρ_R)-space in which the optimal content strategy differs because of ad format. This holds even when $r = 0$ because of the increased consumer distraction ($\hat{c}_s^S < \hat{c}_s^N$). Part (ii) addresses the impact of payment attribution. Increasing the cost of advertising, $r \uparrow$, raises the advertiser's incentive to lead with I under the skippable format. Hence, the gap between Γ^S and Γ^N increases in $r > 0$.

Before moving on to entertaining content, it should be acknowledged that viewers sometime see ads they have seen before. In many cases, the skippable format enables them to avoid useless repeated views. In other cases, they may have forgotten the ad and repetitive views act as beneficial reminders. Within the context of our model, we can think of two different reminder conditions. In one condition, the viewer sees either signal to be reminded of the ad and then anticipates the realization of the second signal without paying attention to it. This condition would not change the viewer's conversion behavior and, therefore, the advertiser's equilibrium strategy described in Propositions 1 and 2 remains unchanged. In the second reminder condition, the viewer is reminded only after she receives the brand information. Obviously, the advertiser leading with R changes nothing about viewer behavior. However, by leading with I the advertiser can gain a strong conversion (with probability $\rho_I\rho_R$) with the viewer not attending to the second part of the ad. This logic applies to both ad formats and raises the advertiser's incentive to lead with I. And further, the skippable format with payment attribution affords the advertiser the possibility to obtain a strong conversion without even paying for the ad, thereby elevating the short-attention-span motive relative to Proposition 2.

4. Using Entertaining Content

In addition to choosing content order, the advertiser can also strategically combat viewer distractions by investing in entertaining content. An amount of entertainment provides *non-informational* benefit e for consumers in the second part of the ad at the cost ke^2. Our conceptualization does not rule out the possibility that the viewer enjoys some of entertainment benefits during the first part of the ad. Moreover, because we model the attention decision only at the end of $t = 1$, the advertiser's choice e merely specifies how much benefit there will be to the viewer during the second part of the ad. Critical for our model, however, is that the viewer observes the value of e during the first part of the ad and that attention is needed in period 2 for the viewer to enjoy the benefit from e. For example, the viewer can infer e based on the plot setup in the first part of the ad and then enjoy the benefit only after finishing the story in the latter part of the ad.[27] Optimizing profits in (3), (4), (6) and (7) with respect to e leads to the equilibrium levels of advertising, $e_{\mathcal{s}}^N$,$e_{\mathcal{s}}^S$, $\mathcal{s} \in \{N, S\}$, whose expressions are provided in Table A1 of the Appendix and order properties given in the following proposition.

Proposition 3 *For any* $V_W \geq 0$, *let* $k > \underline{k} \equiv \frac{\rho_I}{2\bar{c}\widehat{w}}[\rho_R V_S + (1-\rho_R)V_W]$.

(i) *For* $r > 0$, *the non-skippable format leads to strictly more entertainment:* $e_{\mathcal{s}}^N > e_{\mathcal{s}}^S$.

(ii) *Leading with signal R induces weakly more entertainment under both formats: for* $f \in \{N, S\}$, $e_I^f \leq e_R^f$, *with equality if and only if* $V_W = 0$.

The condition on k ensures that, $e < \widehat{w}$, as assumed in Section 3, holds in equilibrium. For part (i), when the payment attribution property is active ($r > 0$), the skippable format offers

[27] It is acknowledged that, without commitment ability, the advertiser has the possibility to renege on a claim in $t = 1$ that a level $e > 0$ for $t = 2$ is forthcoming. Such commitment can be justified on grounds similar to perception formation under umbrella branding (Keller & Lehmann and Almadoss & Jain 2015). For example, suppose that the viewer will immediately have a bad impression ($I = 0$) upon the realization of unfulfilled entertainment.

a lower rate of return to entertainment regardless of the informational content ordering and the availability of a weak conversion ($V_W \geq 0$). Part (ii) shows how the short-attention-span motive interacts with entertainment's return. Recall from Proposition 1, that this motive is active only if the advertiser obtains positive benefit from a weak conversion, $V_W > 0$. In this case, the advertiser has a greater incentive for entertainment under R than under I. Even if $R = 0$ in the first ad period, entertainment can encourage a weak conversion by having the viewer stick around for the second period. However, if $I = 0$ in the first period, then the advertiser gains no benefit from the viewer's continued viewing in the second period. Thus, the short-attention-span motive, which favors leading with I, comes with a lower return on investing in entertainment. This effect disappears when $V_W = 0$; the order of signals does not affect the amount of entertainment because the sole source of advertiser profit is from a strong conversion.

Lastly, we derive several insights from the comparative statics implied by expressions for e_s^f in Table A1 (following References). Entertainment levels are uniformly increasing in $\rho_R \rho_I V_S$, the expected benefit of a strong conversion. An increase in this benefit directly raises the return to entertainment. Entertainment levels are (weakly) decreasing in V_W, generally because the increased benefit from a first-period weak conversion makes a strong conversion incrementally less valuable. An exception to this rule is e_R^N. This exception follows because the advertiser can hope to obtain a weak conversion even when $R = 0$. In this special case, if the viewer is induced to watch the second signal when the first signal is unsuccessful (with probability $1 - \rho_R$), then the advertiser still has a chance at a weak conversion. The assumption $k > \underline{k}$ ($\therefore e < \widehat{w}$) implies this incentive cannot be present under the skippable format. Finally, entertainment levels are uniformly decreasing in $\bar{c}$. An increase in exogenous attention costs reduces the portion of viewers continuing to the second part of the ad, thereby reducing the marginal return of entertainment.

We turn to the advertiser's optimal order of signals and start with the simple case of $V_W = 0$.

Proposition 4 (Optimal Signal Order with Endogenous e) *There exists* $\overline{V} > 0$ *such that for any* $\rho_I > 0$ *and* $V_W \in (0, \bar{V})$, *the following are true.*

(i) *Leading with I is optimal for ρ_R above the indifference curve Γ^f. For $f \in \{N, S\}$, $\pi_I^f > \pi_R^f$, if and only if $\rho_R > \Gamma^f(\rho_I, V_W)$.*

(ii) *Increasing the advertiser's benefit from a weak conversion expands the equilibrium condition for leading with I and to a greater extent for the skippable format: Under the conditions of Proposition 3, the curves Γ^f are decreasing in V_W and $\Gamma^S(\rho_I, V_W) < \Gamma^N(\rho_I, V_W) < \rho_I$ for all $\rho_I > 0$.*

(iii) *Raising either party's benefit from a strong conversion expands the equilibrium condition for leading with R:Both curves Γ^f are increasing in u_S and in V_S.*

(iv) *Increasing savings from payment attribution raises the firm's incentive to lead with I: The curve $\Gamma^S$is decreasing in $r \geq 0$.*

Proposition 4 extends the intuitive mechanisms of Propositions 1 and 2 to the case of endogenous e. Figure 2 can, again, be used for interpretation. Part (i) establishes that the indifference curve interpretation used with fixed e continues to hold. Part (ii) further establishes that the indifference curves for skippable ads lie below that for non-skippable and both below the 45°-line ($\rho_I = \rho_R$). Parts (iii) and (iv) show that the same comparative statics with respect to u_S, V_S, and r apply to the case of endogenous e. Therefore, the equilibrium order of informational content I and R has the same general intuition as described in Section 3.

The unique insight arising from the case of endogenous e requires a deeper look into part (ii) of Proposition 4. Recall that, with fixed e, if $\rho_R \in (\Gamma^S, \Gamma^N)$ the curiosity-inducing motive dominates under the non-skippable format and the short-attention-span for the skippable format. When choosing how to invest in entertainment, however, the advertiser reacts quite differently depending on the ad format. Under the non-skippable format and leading with R, the advertiser seeks a strong conversion and augments the curiosity-inducing motive with a substantial investment in entertainment. But, under the skippable format, the advertiser leading with I is banking on the weak conversion of the short-attention-span motive, which lowers the return on entertainment. Thus, in the regime between the indifference curves (Figure 2), a switch in formats from non-skippable to skippable causes a drastic reduction in entertainment investments. This can be seen also from the expressions in Proposition 3. That is, under the conditions of part (ii) of Proposition 4, $e_R^N = \max\{e_s^f\}$ but $e_I^S = \min\{e_s^f\}$. Entertainment takes a double hit from a switch in formats. First, a switch to the skippable format reduces the return to entertainment because of payment attribution $r > 0$. Second, this switch also induces a change in information order, from R to I, which we know from Proposition 3 to also reduce the return to entertainment. We interpret this and the other formal results from Propositions 1-4 in a practical context in the next section.

5. Interpreting the Results

The model above can be adopted to a decision framework for managers to identify optimal content strategies under various market conditions. Our decision framework is based on marketing contexts characterized on the three traits.

(i) Market Factors. The market in which the advertiser operates is subject to various category factors, such as frequency of consumption, popularity, or the number of alternatives, each of which affects consumers' market options and consequently their inherent interest in learning more from an ad.

(ii) Advertising Goals. Marketers main goal is to inspire action, such as a purchase or donation – a *strong conversion*. But the advertiser can also have sub-goals such as simple brand awareness or follow-on search[28] – a *weak conversion*. The relative value the marketer places on these goals can be a critical factor in seeking content.

(iii) Advertising Formats. Skippable and non-skippable ad formats each imply different levels of distractions for the viewer. Another important feature of skippable ads on digital media is *payment attribution*, the ability of the advertiser to avoid paying for the delivery of an ad that is skipped (Dukes et al. 2022), which affects an advertiser's incentive to induce attention.

Our equilibrium analysis ensures that each content strategy properly assesses the viewer's motivation to pay attention. While there are many factors that influence advertising success, we have focused on the three that we believe are prone to moderate the strategic interaction of content and attention *within a single ad*. Accordingly, we do not generate insights on other strategic aspects of advertising campaigns such as reach, broadness of appeal, and measurement. Also keep in mind that our interpretation of content is indirect and not always an explicit

[28] Joo et al. (2014) show how ad viewing on television can lead to subsequent online search, as measured by the Google metric "Follow-on Search" (Pashkevich et al. 2012).

representation of all advertising elements. Therefore, the interpretations of examples of advertising practice should be viewed as generalizations intended to illustrate the basic principles of our model rather than to offer evidence for our theory.

Our framework is depicted as a decision process in Figure 3. The first step in the process is to assess the existing market factors, which reflect the extent of options for the consumer within the category relative to the popularity of the general category. This step is represented at the top diamond of Figure 3. Consider first the case of $\rho_I < \rho_R$, which depicts a market setting in which there are alternative product options (small ρ_I) in a relatively broad category (large ρ_R). In Figure 3, this corresponds to the left branch off the "Market Setting?" prompt. A curiosity-inducing motive is predominant for informational content and the advertiser leads with the brand-impression signal. By leading with brand impression, the advertiser ensures that whoever responds positively to the brand will be curious to find out about the product category and its relevance. Many brands exhibit this advertising strategy when they introduce new product features or updates. For example, an ad for a discount chain of hotels informs consumers of its modernized décor as well as updated amenities such as free wi-fi, free breakfast, and loyalty programs.[29] The ad begins by introducing the brand to captivate the core consumer's attention (anyone with a positive brand impression) to inform herself of the updated amenities and décor.

Now consider the condition $\rho_I > \rho_R$, which is interpreted as an offering in a market with few available options within the category in a relatively narrow product category. This parametric condition reflects that the brand impression is more likely to be favorable than the relevance of the product itself. While the market may recognize the brand favorably on average, few consumers are in the market for its products. The content strategy in this setting depends on ad goals, which is the second decision diamond in Figure 3 (following References). For a singular ad goal (small V_W/V_S) or large viewer motivation (large u_S), the advertiser structures informational content to appeal to that narrow market by leading with R and inducing those consumers to pursue the seller's information. As with the first case above, the guiding motive is curiosity inducing except that the content order is reversed. This "baiting" strategy can be applied to the practice of clickbait, which is used to generate web traffic from a narrow set of consumers who are keenly interested in knowing the seller. Indeed, much of clickbait banners

[29] One can view the ad here: https://www.youtube.com/watch?v=MPe3LlPfKR4.

contain no brand information and rely on the rare viewer who finds the product or service potentially relevant (small ρ_R) and has a strong motive (large ρ_I) to find out more about the seller.

If ad goals are non-singular (large V_W/V_S) or the viewer is less motivated for conversion (small u_S), then the ad format plays the decisive role. This corresponds to the "Ad Format?" diamond in Figure 3. Under the skippable format, the advertiser is even less inclined to rely only on the consumer's desire to learn more. The guiding motive here is the short-attention span. That is, even though the curiosity inducing motive would suggest leading with R ($\rho_I > \rho_R$), the advertiser would like to ensure at least a weak conversion by leading with I. Of course, some viewers who have a favorable brand impression ($I = 1$) and a low attention cost will not skip, potentially leading to a strong conversion. This situation can occur as an ad in print media, such as in a magazine, in which the advertiser expects the typical reader to continue turning the page, but hopefully obtaining a favorable brand impression, before doing so. Interested readers, of course, can find details of the product's relevance by spending more time and attention to the ad. The short-attention span motive is even more acute for skippable ads on digital media whenever there is payment attribution ($r > 0$). In this way, a weak conversion can occur without the advertiser paying for it.[30]

Under the non-skippable format, the advertiser has a better chance of retaining the viewer for the second part of the ad because of fewer distractions. This means the content strategy can take better advantage of the curiosity inducing motive and lead with R. Furthermore, the advertiser has a better return on entertainment because it has chances to capture both weak and strong conversion upon continued viewing. Even if the viewer finds the product category irrelevant ($R = 0$), entertainment may induce her to continue watching and receive a positive brand impression. In fact, in this equilibrium regime (between curves Γ^N and Γ^S in Figure 2), we see the highest level of investment in entertainment across all regimes. Here one can have in mind complex messages for which the viewer receives little benefit from a conversion (small u_S) and the advertiser cares about both kinds of conversions (large V_W/V_S). Consider, for instance philanthropic messages or public image campaigns where the advertiser cares a lot about a

[30] Google promotes its skippable format for exactly this benefit. See https://www.thinkwithgoogle.com/marketing-strategies/video/creating-youtube-ads-that-break-through-in-a-skippable-world/.

conversion whereas the viewer less so.[31] Also reflective of this type of ad is the, so called, *mystery ad* that appears in video formats, in which the viewer is not initially informed of the advertising brand (Fazio et al. 1992). A mystery ad reveals the product category as well as a compelling narrative to keep the viewer's attention. Imagine, for instance, a performance luxury car ad mixing details of the automobile and a small jet aircraft accelerating on the runway and lifting off, mentioning the car brand only at the very end.[32] Table 3 (following References) summarizes the above discussion with some examples.

Finally, we interpret one more finding regarding entertaining content across ad formats. One may have intuited that the advertiser should invest more, not less, in combating a lack of viewer attention with skippable ads. But, from Proposition 3, we observe that viewers have more distractions with skippable ads, which makes it harder to keep their attention and reduces the marginal return of investment in entertainment. While the formal result suggests that skippable ads will never have more entertaining content than non-skippable, we acknowledge the reality can be different in some cases. Indeed, there are many examples of highly entertaining skippable ads and even glossy magazine ads involving expensive production. However, the mechanism offered in our model speaks to an aggregate level trend seen in digital advertising. One wisdom that has been expressed by industry observers that digital ads are of lower copy quality than non-digital ads (O'Guinn 2009). Indeed, ads appearing during spectacle events on television (e.g., Super Bowl) are considerably more costly to produce than the typical pre-roll ad on YouTube. While there are likely multiple reasons for this distinction, our results suggest that short-attention span motives and payment attribution are also contributing factors.

An important caveat to our representative consumer model is the assumption that an ad's reach is one. A literal interpretation of this assumption is that targeting is perfect and all consumers in the market have the same parametric values. Obviously, this is a restrictive setting and reflects very few actual markets. However, it can provide a starting point for more realistic settings. We briefly depart from the framework in two directions – reach less than one and a market with consumer heterogeneity.

---

[31] For example, as part of a campaign to strengthen its public image, Walmart aired a television ad showing families and military communities "Giving Thanks" at a holiday meal before revealing themselves as the advertiser. See https://www.youtube.com/watch?v=8mBixdPUG74.
[32] This description is from a real ad seen here https://www.youtube.com/watch?v=sV5MwVYQwS8.

When reach < 1, it means that an ad reaches only a subset of the entire market. A convenient feature of our monopoly model is the clean distinction this departure has on informational and non-informational content. Specifically, as long as there is no consumer heterogeneity, any partial reach has direct implications for entertainment level, but not on information order. To see this, note in profit equations (3), (4), (6), and (7), a reach of $\phi < 1$ can be represented by the proportion in front of revenue terms, but not entertainment costs ke^2. The corresponding optimal entertainment levels are subsequently discounted by the same factor ϕ. As the portion of consumers in the market decreases, so too does the return on the investment in entertainment. And, because the advertiser's choice of content order has no cost differential, the optimal order is determined by a comparison of per person expected revenue, or equations (5) and (8). We qualify that the above logic presumes a monopoly advertiser. There is no guarantee that our results hold when there are competing advertisers.

Incorporating consumer heterogeneity is more involved. The additional challenge is in assessing the distribution of the key market parameters ρ_R, ρ_I, and u_S, as well as some reasonable assumptions on the attention cost distribution c. An easy approach is to proxy the market parameters with segment averages and an estimate of the marginal attention cost $\hat{c}/\bar{c}$, through, say, a conjoint survey. The accuracy of this approach obviously depends on the degree of variance of market parameters. In the case of wide population variance, simulations on (3) and (4) for non-skippable, or (6) and (7) for skippable, must be done at the individual level and then aggregated to determine the optimal content order.

6. Additional Considerations

Our setting excludes several practical elements that can arise in ad content decisions. In this section, we introduce three such elements to gain additional insights and explore the boundaries of our basic model.

6.1 Ad Annoyance

Ad annoyance and other ill-feelings from an ad's interruption are already subsumed within the notion of attention cost in the main model (Todri et al. 2020). However, there is evidence that the incurrence of annoyance arises when the ad is irrelevant (Aaker and Bruzzone 1985).[33] This

[33] We thank the review team for suggesting this possibility.

raises the possibility that attention cost c can depend on the outcome of the first informative signal, I or R, which has been heretofore ignored. Therefore, in this section, we consider the possibility that a positive signal in period 1 can attenuate the viewer's attention cost in period 2. We take up this possibility here to better understand how ad annoyance can affect the advertiser's choice of content.

To implement this consideration, suppose a viewer's attention cost c is reduced if the viewer discovers some relevance to the category or strikes a positive impression in period 1. Specifically, we introduce an *annoyance attenuation parameter* $\lambda \leq 1$ which reduces the viewer's attention cost from c to λc if and only if the first signal (I or R) is favorable. An unfavorable signal maintains attention cost c as previously analyzed. Therefore, the impact of annoyance associated with an unfavorable signal is larger for lower values of λ. We also assume that λ is sufficiently high to maintain the interior attention cost cutoffs between 0 and $\bar{c}$. Note that setting $\lambda = 1$ recovers our baseline model and, by continuity, previous results carry through qualitatively when $\lambda \uparrow 1$.

The new insights from this extension can be exposed using the basic intuition of the baseline model and without the need for analytic details. We start with informative ad content and the ordering of I and R. If $V_W = 0$, then changes in λ imply no change in the results because the curiosity-inducing motive exclusively drives the ordering via the ordering of the ρ's. We now suppose $V_W > 0$ and argue that a reduction in λ shifts Γ^f upward. When attention cost is reduced by a favorable first signal, viewers are more likely to continue watching the second signal. This enhances advertiser's ability to focus on a strong conversion over a weak conversion, making short-attention motive incentive marginally less attractive, thereby rotating the curve Γ^f upward toward the 45°-line.

Next, we ask how annoyance attenuation moderates the impact of ad format. Recall from Proposition 2 that the skippable format favors I over R, as noted in the wedge between the iso-profit curves of Figure 2. In general, the short-attention span motive is stronger under the skippable format than under the non-skippable, with the gap representing the region in which the format matters for the choice of I versus R. As annoyance attenuation increases, the short-attention span motive becomes incrementally less attractive to the advertiser relative to the curiosity-inducing motive, as argued above. So, while both curves Γ^N and Γ^S shift upward when

$\lambda \downarrow$, the gap between them shrinks. That is, skippable format's favoring of I is reduced for a reduction in λ.

Finally, we discuss impact of λ on the advertiser's incentive to invest in entertaining content. It is tempting to intuit that annoyance attenuation substitutes for entertainment, due to the viewer's additional compensation $(1-\lambda)c$. However, this added compensation actually complements the return on entertainment because, a decrease in λ raises the threshold $\hat{c}$ upon favorable information. Therefore, all else equal, more viewers attend to the second part of the ad thereby raising the advertiser's return on investment e.

In summary, the incentives for both types of content are affected in the presence of ad annoyance, when attenuated due to favorable match values $(I, R = 1)$. For informational content, more ad annoyance attenuation *raises* the scope for the curiosity-inducing motive in leading with the smaller of the ρ's. For non-informational content, greater ad annoyance attenuation raises the optimal level of entertainment.

6.2 Platform Incentives

Many advertisements are delivered through an intermediary, such as a commercial content platform. In this section, we discuss how the platform's own profit incentive can affect the advertiser's content decisions implied in our model. In doing so, we refrain from any conclusion about what is optimal for the platform because we do not model how the price of ads r is determined.[34]

An important aspect of any platform is in predicting the tastes of its viewers so that advertisers can better target ads. A key principle in targeting is in segmenting consumers so that, within each segment, consumers have similar traits. Recall that our representative viewer model treats all viewers the same. As such, our model is best suited for situations in which viewers can be assumed to have similar ρ's and u_S's, such as in highly targeted environments like online platforms.

Another important consideration for the platform is in attracting and retaining viewers. This means that the platform should assess how well the advertiser's decisions on ad content are aligned with viewer's preferences. We have ignored this alignment until now. In terms of

[34] For this reason, our framework is unable to explain platform decisions on ad format, say between non-skippable and skippable, which is an important area for future research.

informational content, we find (in separate analysis) that the viewer's preference for signal order can be opposite of the advertiser's. In fact, viewers with intermediate attention costs have a higher expected utility when the advertiser starts with the *stronger* signal whereas the advertiser, acting on the curiosity-inducing motive, prefers to start with the weaker signal. For these viewers, a stronger starting signal is more diagnostic of the viewer's expected benefit from paying additional attention.[35] Furthermore, no viewer is strictly better off from leading with the weaker signal. Interestingly, the platform's use of the skippable format can realign advertiser preferences with the viewer's. To see this, suppose $V_W > 0$ and (ρ_I, ρ_R) within the wedge created by iso-profit curves Γ^S and Γ^N (referring to Figure 2). In this case, under non-skippable ads, the advertiser choses R in equilibrium because it is the weaker signal ($\rho_R < \rho_I$) yet the viewer prefers the opposite order. But now, if the platform implements the skippable ad format, the advertiser optimally takes advantage of the short-attention span motive by choosing I, which is the viewer's preferred order.

Relatedly, the platform should care about the quality of ads shown to viewers. According to Proposition 3, the platform's use of the skippable ad format leads to lower equilibrium levels of e relative to non-skippable. This has consequences for digital content platforms and can help us understand how skippable ads are marketed. Because ad quality affects a platform's attractiveness for viewers, it wants to encourage advertisers to produce good quality ads. This observation could explain why YouTube and Facebook help and advise advertisers to make skippable ads compelling to watch (i.e., make e bigger).[36] Relatedly, our results can also explain the value proposition of the skippable ad format marketed by these and other digital content platforms. These platforms advise advertisers to put their brand identifier early in the ad, so that if the viewer skips, she still receives the brand impression. Google, for instance, even tells advertisers that this is a cheap means to acquire brand awareness without paying the full ad price.[37]

---

[35] For instance, whenever $V_W = 0$ and $\rho_R < \rho_I$, the advertiser leads with R in equilibrium, but any viewer with $c \in (\tilde{c}_R^N, \hat{c}_R^N)$ has strictly more expected (in period 0) utility when the advertiser leads with I for non-skippable ads. Details in the Supplemental Appendix.

[36] Facebook, for instance, provides online courses on improving video ads to make them more watchable (https://www.facebook.com/business/ads/video-ad-format). Similarly, YouTube helps advertisers create video ads to improve their quality (https://www.youtube.com/ads/making-a-video-ad).

[37] Google's advises advertisers using the skippable format on YouTube to display the brand and logo within the first 5 seconds in order to achieve brand awareness even if the viewer skips. Google's advice is based on internal

6.3 Complementary Roles of Advertising

Often the content of the ad itself has a direct effect on the consumer's utility. Under this *complementary view* of advertising (Becker and Murphy 1993), narratives or personalities that appear in the ad support the consumer's perception of the product itself. For example, an amateur tennis player could derive extra benefit from using a particular racquet simply because the commercial featured his favorite tennis pro. With such a possibility, it is interesting to explore the implications for ad content and attention. In this section, we ask: How does the consideration of the complementary view of advertising affect ad content?

To investigate this setting, we modify our model by augmenting the utility the consumer obtains from a strong conversion. Specifically, redefine u_S to be an increasing function of e. This can be interpreted as the cost of a celebrity endorsement. The simplest formulation is a linear one, $\hat{u}_S = u_S + \alpha \cdot e$, with $\alpha > 0$. Obviously, this has no bearing on the model with fixed e. Henceforth, we consider only endogenous e.

As we see from the expression in Table A2, replacing u_S with $\hat{u}_S$ adds a second vehicle for entertainment to combat inattention. Specifically, a marginal increase in e directly raise the direct utility of watching the ad, and it also gives the viewer more motivation for a strong conversion. As a result, the equilibrium levels of entertainment are increasing in α for each format f and content order I/R. Furthermore, the expressions for $e_{I/R}^{f}$ are restored to those in Proposition 3 when $\alpha = 0$.

The impact of the complementary view of advertising on informational content is less obvious. However, a full equilibrium analysis is not necessary to glean the main insight. First note that, with a larger investment e, the corresponding increase in $\hat{u}_S$ can alter the differential motives for curiosity and short-attention spans. The curiosity inducing motive generates attention by exploiting the viewer's willingness for a strong conversion as measured by $\hat{u}_S$. By contrast, the short-attention span motive encourages the advertiser to settle for a weak conversion, as measured by $V_W > 0$, which is independent of e. Therefore, with the complementary view of advertising at work ($\alpha > 0$), the scope for invoking the curiosity-inducing motive is larger

company research available from the authors. See also https://www.thinkwithgoogle.com/marketing-strategies/video/creating-youtube-ads-that-break-through-in-a-skippable-world/.

relative to the short-attention span. From the advertiser's perspective, the complementary role of advertising diminishes the value of leading with I for short-attention spans and enhances the curiosity-inducing motive when $\rho_R < \rho_I$ (the lower-right portion of Figure 2). We imagine, in our earlier example, the amateur tennis player attending to the entire ad to see the specific racquet his favorite pro is using. All else equal, this insight suggests that if a product that relies on expensive celebrities or personas as part of the brand, then the short-attention-span motive is relatively less useful.

7. Conclusion

A lot of work in marketing has been devoted to understanding the factors of advertising's success and how that success depends on the consumer's attention to an ad. But little research has studied the consumer's economic tradeoffs when deciding whether to pay attention to its content. This paper highlighted the importance of anticipating the consumer's inherent motivation to attend to ad content against their tendency to succumb to distraction in the digital age. Using an equilibrium approach, we examined how informational and non-informational elements of the ad should be structured to manage a consumer's attention while simultaneously meeting communication goals. Such an approach allowed us to identify content strategies that manage a consumer's attention to an ad in a variety of market settings.

Our analysis uncovered two strategic mechanisms for structuring informational ad content. The first mechanism is to induce curiosity – making the viewer motivated to learn more from watching the entire ad. We showed how curiosity induction leads to a key result that, to maximize the viewer's motivation, the advertiser should lead with the weaker of the two information signals. The second mechanism is to manage short-attention-spans – achieving impact even while expecting her to eventually become distracted before the ad is over. As we showed, the return to investing in non-informative, entertaining, content is reduced under the skippable ad format, a format used in many digital contexts.

Admittedly, many contexts have not been considered in this study. Advertising dynamics over time and media are clearly important factors of ad effectiveness. For instance, wear-out, carry-over, scheduling, and multi-media effects were not incorporated in this model. Furthermore, it is likely that many of these aspects moderate a consumer's incentive for attention to ad content. However, incorporating too many of these features would have prevented us from communicating some of the basic mechanisms of ad consumption (e.g., curiosity inducing and short-attention span motives). Future research can hopefully build off and modify this framework to further improve our collective understanding of ad consumption.

Another important caveat to our game-theoretic methodology is the fact that we have treated ad content exclusively in terms of the costs and benefits of attention. While this permitted us to specify a viewer's utility and assess decisions in the framework of utility maximization, it put limitations on other important elements of effective ads. Indeed, advertisers develop ads with other factors in mind, particularly related to emotional appeals, sensory stimulation, or other affective elements. These other factors are likely to have important consequences for viewer attention as well and we have left them unexplored. Our modest hope with this work is to provoke the notion that an assessment of the viewer's individual incentive to pay attention is key to effective advertising in the digital age.

Appendix

Analyzing the Consumer's Attention Decision

Decision trees for each of the I and R scenarios under the skippable format can be obtained by substituting the word "Ignore" with "Skip" and in Figure A1 and A2, respectively. And, within that "skip" branch, replace the viewer's payoff δw with w. The determination of viewers' attention choices follows the same process as with the non-skippable ad format and with the corresponding thresholds in Table A1 (following References).

With the aid of the decision trees and the arguments in Section 2, the viewer's decisions are defined by the attention cost cutoffs given in Lemma A1.

Lemma A1 (Viewer Outcomes for Each Content Strategy) *Assume* $e < \widehat{w}$.

(i) For a favorable first signal $\mathcal{s} \in \{I, R\}$, $\mathcal{s}' \neq \mathcal{s}$, $\hat{c}_{\mathcal{s}}^{N} = \rho_{\mathcal{s}'} u_S + e$, $\hat{c}_{\mathcal{s}}^{S} = \rho_{\mathcal{s}'} u_S + e - \widehat{w}$. *All thresholds* $\hat{c}_{\mathcal{s}}^{f} \in (0, \bar{c})$ *for* $\rho_{\mathcal{s}} u_S + e < \bar{c}$ *under non-skippable ads (N) and for* $0 \leq \rho_{\mathcal{s}} u_S + e - \widehat{w} < \bar{c}$, *under skippable ads (S).*

(ii) For an unfavorable first signal $\mathcal{s} \in \{I, R\}$, $\tilde{c}_R^N = e$, $c_I^N = \tilde{c}_{\mathcal{s}}^S = 0$.

Proof of Lemma A1

The case of non-skippable ads was derived in the main text. The remainder of the proof is devoted to the case of skippable ads.

Suppose the advertiser leads with I. If the viewer pays attention $c > 0$ in period 2, then her utility is

$$U_I^S(c) = \begin{cases} \rho_R u_S + (e - c) & \text{if } I = 1 \\ e - c & \text{otherwise.} \end{cases}$$

If she does not pay attention to the second part of the ad, she can enjoy the desired content right away and receive additional payoff $\widehat{w} > e$, which is the condition of the lemma. Therefore, it is optimal for all viewers to pay no attention when $I = 0$. If $I = 1$, then the threshold viewer $\hat{c}_I^S$ is indifferent to attending to the rest of the ad: $\rho_R u_S + (e - \hat{c}_I^S) = \widehat{w}$, which implies the expression given in Table A1.

When the advertiser leads with R, the same argument above, with subscripts I and R switched, gives the expression for the threshold $\hat{c}_R^N$ as given in Table A1.

Next, we establish the conversion probabilities as claimed. First, suppose that the advertiser leads with I under non-skippable ads. With probability ρ_I, the first signal is favorable ($I = 1$). Viewers with $c \in [0, \hat{c}_I^N]$ (of fraction $\hat{c}_I^N/\bar{c}$) will pay attention to the second signal. With a further probability ρ_R, the second signal is also favorable ($R = 1$) with strong conversion. In the remaining probability $1 - \rho_R$, the second signal is unfavorable so there is only weak conversion. In addition, viewers with $c \in (\hat{c}_I^N, \bar{c}]$ will not wait for the second signal, generating weak conversion only. Moving back to the first signal, if it is unfavorable ($I = 0$), then neither strong nor weak conversion is possible. Thus, the attention threshold is irrelevant from the advertiser's perspective.

Now suppose that the advertiser, again under non-skippable ads, leads with R. With probability ρ_R, the first signal is favorable ($R = 1$). For weak conversion to be possible, $I = 1$ is required. Combined with $R = 1$, any conversion must be a strong conversion (no weak conversion is possible). This occurs for viewers with $c \in [0, \hat{c}_R^N]$ (of fraction $\hat{c}_R^N/\bar{c}$) who pay attention to the second signal, and with further probability ρ_I the second signal is also favorable ($I = 1$). Now suppose that the first signal is unfavorable ($R = 0$) which means that strong conversion is out. Viewers with $c \in [0, \tilde{c}_R^N]$ (of fraction $\frac{\tilde{c}_R^N}{\bar{c}}$) will pay attention to the second signal. With a further probability ρ_R, the second signal is favorable ($R = 1$) which leads to weak conversion.

The case of skippable ads with favorable first signal are quite similar to the counterparts in the non-skippable ads, except that the attention thresholds change. However, if the first signal is unfavorable, no viewer will wait for the second signal. Thus there is no strong or weak conversion. The rest of the conversion probabilities can be derived similarly. ■

Proposition A1 (Optimal Signal Order with Exogenous e) *Suppose $e < \hat{w}$, e fixed, and $\rho_I \rho_R V_S > r(\hat{w}/u_S)$. For some $\bar{V} > 0$ and for each $f \in \{N, S\}$, there exists $\Gamma^f: [0,1] \times [0, \bar{V}) \to \mathbb{R}$ such that $\Delta^f = 0$ whenever $\rho_R = \Gamma^f(\rho_I, V_W)$. Furthermore, for $\rho_I \in [0,1]$ there exists $\xi > 0$, such that whenever $V_W \in (0, \bar{V})$ and $|\rho_R - \Gamma^f(\rho_I, V_W)| < \xi$, $\Gamma^f(\rho_I, V_W)$ has the following properties.*

(iv) $\pi_I^f > (<) \pi_R^f$, *whenever* $\rho_R > (<) \Gamma^f(\rho_I, V_W)$;

(v) $\Gamma^f(\rho_I, V_W)$ *is the 45°-line for* $V_W = 0$ *and is decreasing in* $\frac{V_W}{u_S V_S} > 0$.

Proof of Proposition A1

Part (i) is established with by assessing the sign of the derivatives

$$\frac{\partial \Delta^N}{\partial \rho_R} = 2\rho_R u_S (V_S - V_W) - \rho_I u_S V_S, \quad \text{and}$$

$$\frac{\partial \Delta^S}{\partial \rho_R} = (\rho_I u_S V_S - r\widehat{w})(\rho_R - \rho_I) + (\rho_R \rho_I u_S V_S - r\widehat{w}) + \rho_I(-2\rho_R u_S + \widehat{w} - e)V_W,$$

both of which are guaranteed positive as long as $|\rho_R - \rho_R| < \xi_1$ and $V_W \in (0, \bar{V}_1)$, for some $\xi_1 > 0$ and $\bar{V} > 0$ sufficiently small. Part (ii) is established assigning the signs of the partial derivatives of Γ^f with respect to V_W, u_S, and V_S, each of which are derived from the Implicit Function Theorem where, for each $f \in \{N, S\}$ and $x \in \{V_W, u_S, V_S\}$

$$\frac{\partial \Gamma^f}{\partial x} = -\frac{\partial \Delta^f / \partial x}{\partial \Delta^f / \partial \rho_R},$$

the denominator of which is positive for both $f \in \{N, S\}$ by part (i) when $V_W \in (0, \bar{V})$ and $|\rho_R - \rho_I| < \xi_1$. Starting with non-skippable ($f = N$), the corresponding numerators

$$\frac{\partial \Delta^N}{\partial V_W} = \bar{c} - \rho_R^2 u_S - e > 0,$$
$$\frac{\partial \Delta^N}{\partial u_S} = \rho_R V_S(\rho_R - \rho_I) - \rho_R^2 V_W < 0,$$
$$\frac{\partial \Delta^N}{\partial V_S} = \rho_R u_S(\rho_R - \rho_I) < 0,$$

have the indicated signs for any $\rho_I \in [0,1]$ when $|\rho_R - \Gamma^N(\rho_I, V_W)| < \xi_2$, for some $\xi_2 > 0$. Turning to skippable ads ($f = S$), the corresponding numerators

$$\frac{\partial \Delta^S}{\partial V_W} = -\rho_I[\bar{c} - \rho_R(\rho_R u_S - \widehat{w} - e)] > 0,$$
$$\frac{\partial \Delta^S}{\partial u_S} = \rho_R \rho_I V_S(\rho_R - \rho_I) + \rho_I(-\rho_R^2)V_W < 0$$
$$\frac{\partial \Delta^N}{\partial V_S} = \rho_R u_S(\rho_R - \rho_I) < 0,$$

have the indicated signs for $\rho_I \in [0,1]$ when $V_W \in (0, \bar{V})$ and $|\rho_R - \Gamma^N(\rho_I, V_W)| < \xi_3$, for some $\xi_3 > 0$. The claimed comparative statics in part (ii) follow from for $\xi \equiv \min\{\xi_1, \xi_2, \xi_3\}$. ■

Proof of Proposition 1

This proposition is a special case of Proposition A1.

Proof of Proposition 2

Conditions (5) and (8) imply that $\Delta^S - \Delta^N = \frac{r\widehat{w}}{\bar{c}}(\rho_I - \rho_R) + \frac{\rho_I}{\bar{c}}[\rho_R\widehat{w} + (1-\rho_R)e]V_W$, which is positive if $r(\rho_I - \rho_R) \geq 0$ and $V_W > 0$. In other words, in the region specified in the conditions of Proposition 1, we have $\Delta^S - \Delta^N > 0$, which holds that $\Delta^N = 0 \Rightarrow \Delta^S > 0$. That is, along the curve $\rho_R = \Gamma^N(\rho_I, V_W)$, which is below ρ_I ($\because V_W > 0$), the advertiser's optimal starting signal is I under the skippable format. By Proposition 1 (i), therefore, the advertiser's signal iso-profit curve under the skippable format must lie below its iso-profit curve under the non-skippable: $\Gamma^N(\rho_I, V_W) > \Gamma^S(\rho_I, V_W)$. Hence, part (i) is established.

To show part (ii), it should be obvious to see that the advertiser's optimal signal order is independent of r. Thus, Γ^N does not shift with changes in r. To establish the shift in Γ^S as claimed, again employ the Intermediate Value Theorem.

$$\frac{\partial\Gamma^S}{\partial r} = -\frac{\partial\Delta^S/\partial r}{\partial\Delta^S/\partial\rho_R}.$$

The denominator above was shown to be positive in the proof of Proposition 1(i). The numerator is the derivative of (8), or $-\widehat{w}(\rho_R - \rho_I)$, which is positive for $\rho_R < \rho_I$, the region below the 45$^{\text{o}}$-line implied by the conditions in Proposition 1. Because both numerator and denominator are positive, the overall derivative with respect to r is negative. ■

Proof of Proposition 3

The four expressions in Table A2 (following References) are solutions to the first-order conditions for optimizing the profits given in (3), (4), (6), and (7) respectively. Each of these solutions is a maximizer because, as is directly verified, the profits are globally concave in e: $\partial^2\pi_{s}^{f}/\partial e^2 = -2k < 0, f \in \{N, S\}, s \in \{I, R\}$. Results (i) through (iii) are obvious using the expressions in Table A2. ■

Proposition A2 *For some* $\bar{V} > 0$ *and for each* $f \in \{N, S\}$*, there exists* $\Gamma^f: [0,1] \times [0, \bar{V}) \to \mathbb{R}$ *such that* $\Delta^f = 0$ *whenever* $\rho_R = \Gamma^f(\rho_I, V_W)$*. Furthermore, for* $\rho_I \in [0,1]$ *there exists* $\xi > 0$*, such that whenever* $V_W \in (0, \bar{V})$ *and* $|\rho_R - \Gamma^f(\rho_I, V_W)| < \xi$*,* $\Gamma^f(\rho_I, V_W)$ *has the following properties.*

(i) $\pi_R^f < (>) \pi_I^f$ *if and only if* $\rho_R > (<) \Gamma^f(\rho_I, V_W)$.

(ii) $\Gamma^f(\rho_I, V_W)$ *is strictly increasing in* ρ_I *with* $\Gamma^f(0, V_W) = 0$.

(iii) *There exists a value* k^f *such that* $\Gamma^f(\rho_I, V_W)$ *is increasing (decreasing) in* V_W *whenever* $k < (>) k^f$.

$k^N = \frac{\rho_I \rho_R V_S}{2\bar{c}(\bar{c} - \rho_R^2 u_S)}$ *and* $k^S = \frac{\rho_I \rho_R(\rho_R V_S - r)}{2\bar{c}[\bar{c} - \rho_R(\rho_R u_S - \hat{w})]}$.

(iv) *These thresholds satisfy* $k^S < k^N$*, so that* $(k^S, k^N) \neq \emptyset$.

Proof of Proposition A2

Recall that $\Delta^f \equiv \pi_I^f - \pi_R^f, f \in \{N, S\}$. We first define the advertiser's indifference condition, $\Delta^f(\rho_R, \rho_I) = 0$, $f = N, S$, as a function in (ρ_I, ρ_R)-space. For any given $V_W \geq 0$, let $\Gamma^f(\rho_I, V_W)$ be the value of ρ_R such that $\Delta^f(\Gamma(\rho_I, V_W), \rho_I) = 0$.

(i) Fix $\rho_I \in (0,1)$ and consider the partial derivatives $\partial\Delta^f/\partial\rho_R$, $f = N, S$. Substituting in to (3), (4), (6), and (7) with the expressions for $\hat{c}_{I/R}^f$ and $e_{I/R}^f$ given in Tables A1 and A2, respectively, we can determine that

$$\lim_{V_W \downarrow 0} \lim_{\rho_R \to \rho_I} \frac{\partial \Delta^N}{\partial \rho_R} = \frac{V_S \rho_I^2 u_S}{\bar{c}}, \tag{A1}$$

and

$$\lim_{V_W \downarrow 0} \lim_{\rho_R \to \rho_I} \frac{\partial \Delta^S}{\partial \rho_R} = \frac{\rho_I^2 V_S u_S - \hat{w} r}{\bar{c}} + \frac{\rho_I r}{2k\bar{c}^2}(\rho_I V_S - r). \tag{A.2}$$

The limit in (A1) is clearly positive. For the limit in (A2), note that the condition in Proposition 1, $\rho_I \rho_R V_S > r\left(\frac{\hat{w}}{u_S}\right)$, implies that the sum in the numerator of the first term is positive and that $e_R^S \geq 0$ implies the second term is non-negative. Combined, we have $\partial\Delta^f/\partial\rho_R > 0$, $f \in \{N, S\}$. By virtue of continuity of Δ^f in V_W and $\rho_{I/R}$, $\Delta^f > (<) 0$ if and only if $\rho_R > (<) \Gamma^f(\rho_I, V_W)$.

That is, there exists values $\bar{V} > 0$ and $\xi > 0$ such that $\xi > \rho_R - \Gamma(\rho_I, V_W) > 0$ implies $\Delta^f = \pi_I^f - \pi_R^f > 0$ and $-\xi < \rho_R - \Gamma(\rho_I, V_W) < 0$ implies $\Delta^f = \pi_I^f - \pi_R^f < 0$, for all $V_W \in (0, \bar{V})$ and $\rho_I \in (0,1)$. This completes the proof of part (i).

(ii) We employ the Implicit Function Theorem to determine the derivative of $\Gamma(\rho_I; V_W)$:

$$\frac{d\Gamma^f(\rho_I;V_W)}{d\rho_I} = -\left(\frac{\partial\Delta^f}{\partial\rho_I}\right)\left(\frac{\partial\Delta^f}{\partial\rho_R}\right)^{-1},$$

the denominator of which was calculated in the limit in (i) above. Substituting in to (3), (4), (5), and (6) with the expressions for $\hat{c}_s^f$ and e_s^f given in Tables A1 and A2, respectively, we can determine that

$$\lim_{V_W\downarrow 0}\lim_{\rho_R\to\rho_I}\frac{\partial\Delta^f}{\partial\rho_I} = -\lim_{V_W\downarrow 0}\lim_{\rho_R\to\rho_I}\frac{\partial\Delta^f}{\partial\rho_R},$$

for $f \in \{N, S\}$. Then $\frac{d\Gamma^f(\rho_I;V_W)}{d\rho_I} \to 1$ as $V_W \downarrow 0$ and $\rho_R \to \rho_I$. This establishes the claim that for any $V_W \in (0, \bar{V})$, with $|\rho_I - \rho_R| < \xi$, the indifference curve $\Gamma(\rho_I, V_W)$ is increasing in ρ_I. Further, it is directly verified that $\Gamma^f(0; V_W) = 0$.

(iii) Again, we employ the Implicit Function Theorem to establish our claim. Specifically, the following needs to be established:

$$\frac{\partial\Gamma^f(\rho_I;V_W)}{\partial V_W} = -\left(\frac{\partial\Delta^f}{\partial V_W}\right)\left(\frac{\partial\Delta^f}{\partial\rho_R}\right)^{-1}\begin{cases} > 0 & k < k^f \\ < 0 & k > k^f. \end{cases}$$

By part (i), $\left(\frac{\partial\Delta^f}{\partial\rho_R}\right)^{-1} > 0$ so it remains only to show that

$$-\left(\frac{\partial\Delta^f}{\partial V_W}\right) > 0 \Leftrightarrow k < k^f. \tag{A3}$$

Substituting in to (3), (4), (6), and (7) with the expressions for $\hat{c}_{I/R}^f$ and $e_{I/R}^f$ given in Lemma A1 and Proposition 3, respectively, significant algebraic manipulations can determine that the condition in (A3) is satisfied for the expressions k^f given in the statement of the proposition.

(iv) For $r, \hat{w} > 0$, we can directly verify that $k^N > k^S$:

$$k^N = \frac{\rho_I\rho_R V_S}{2\bar{c}(\bar{c}-\rho_R^2 u_S)} > \frac{\rho_I\rho_R(\rho_R V_S - r)}{2\bar{c}(\bar{c}-\rho_R^2 u_S)} > \frac{\rho_I\rho_R(\rho_R V_S - r)}{2\bar{c}[\bar{c}-\rho_R(\rho_R u_S - \widehat{w})]} = k^S. \blacksquare$$

Proof of Proposition 4

The claims in (i) follow from Proposition A2. For part (ii), we show that, for any $k \geq \underline{k}$, $k > k^N (> k^S)$. Once this is established, Proposition A2 (iii) gives the result. Henceforth, we need only establish that the lower bound for the proposition's condition on entertainment costs: $\underline{k} \geq k^N$. This lower bound is specified as the minimum level of k such that $\max\{e_s^f\} \leq \widehat{w}$. Note that, as $V_W \downarrow 0$, $\underline{k} \rightarrow \frac{\rho_I\rho_R V_S}{2\bar{c}\widehat{w}}$. This limit is bounded below by k^N if and only if $\bar{c} > \rho_R^2 u_S - \widehat{w}$. Assumption 1 (i) gives us $\rho_R u_S - \widehat{w} < \bar{c}$, which is clearly sufficient for our condition because $\rho_R \in (0,1)$. Further note that, for $k = k^N$, $\Gamma^N(\rho_I, V_W)$ is invariant to changes in V_W but $\Gamma^S(\rho_I, V_W)$ is decreasing because $k > k^S$. Therefore, for $k > \underline{k}$, $\Gamma^S(\rho_I, V_W) < \Gamma^N(\rho_I, V_W) < \rho_I$ for all $\rho_I > 0$ and $V_W \in (0, \bar{V})$.

For parts (iii) and (iv), we use the Implicit Function Theorem to calculate the signs of the appropriate derivatives of the indifference curve Γ^f, $f \in \{N, S\}$:

$$\frac{d\Gamma^f(\rho_I; V_W)}{dx} = -\left(\frac{\partial\Delta^f}{\partial x}\right)\left(\frac{\partial\Delta^f}{\partial\rho_R}\right)^{-1}, \text{ for } x \in \{u_S, V_S, r\}, \tag{A4}$$

ss $V_W \downarrow 0$. From the proof of Proposition A2 we know that $\lim(\partial\Delta^f/\partial\rho_R) > 0$. Therefore, the sign of the overall derivative in (A4) is opposite the sign of $\lim(\partial\Delta^f/\partial x)$. Starting with $x \in \{u_S, V_S\}$:

$$\frac{\partial\Delta^N}{\partial u_S} = \rho_I\rho_R(\rho_R - \rho_I)V_S - \rho_R^2 V_W \rightarrow \rho_I\rho_R(\rho_R - \rho_I)V_S < 0,$$

$$\frac{\partial\Delta^S}{\partial u_S} = \rho_I\rho_R(\rho_R - \rho_I)V_S + [(1-\rho_R)\rho_R\rho_I - \rho_I\rho_R]V_W \rightarrow \rho_I\rho_R(\rho_R - \rho_I)V_S < 0,$$

$$\begin{aligned}\frac{\partial\Delta^N}{\partial V_S} = \rho_I\rho_R &\left[u_S(\rho_R - \rho_I) + (e_I^N - e_R^N) + \left(\frac{\partial e_I^N}{\partial V_S} - \frac{\partial e_R^N}{\partial V_S}\right)V_S\right] \\ &+ \left[-\rho_I\rho_R\left(\frac{\partial e_I^N}{\partial V_S} - \frac{\partial e_R^N}{\partial V_S}\right) - \rho_I\frac{\partial e_R^N}{\partial V_S}\right]V_L \\ &-2k\left(e_I^N\frac{\partial e_I^N}{\partial V_S} - e_R^N\frac{\partial e_R^N}{\partial V_S}\right) \rightarrow \rho_I\rho_R(\rho_R - \rho_I)u_S < 0;\end{aligned}$$

$$\frac{\partial \Delta^S}{\partial V_S} = \rho_I \left\{ \rho_R \left(\frac{\partial e_I^S}{\partial V_S} - \frac{\partial e_R^S}{\partial V_S} \right) V_S + \rho_R [u_S(\rho_R - \rho_I) + (e_I^N - e_R^N)] + \left[(1-\rho_R) \frac{\partial e_I^S}{\partial V_S} - \frac{\partial e_R^S}{\partial V_S} \right] V_W \right\}$$

$$-r \left(\rho_I \frac{\partial e_I^S}{\partial V_S} - \rho_R \frac{\partial e_R^S}{\partial V_S} \right) - 2k \left(e_I^S \frac{\partial e_I^S}{\partial V_S} - e_R^S \frac{\partial e_R^S}{\partial V_S} \right) \to \rho_I \rho_R (\rho_R - \rho_I) \left(u_S + \frac{r}{2k} \right) < 0;$$

where all inequalities employ the fact that $\rho_R < \rho_I$, which holds for $V_W > 0$ in the neighborhood of $V_W = 0$. Hence, Γ^N is increasing in u_S and V_S. For the case of r, a bit of algebra establishes the following expression for the limit $\partial \Delta^S / \partial r \to (\rho_R - \rho_I)(e_R^S - \widehat{w})$, which is positive by the maintained assumption $e_R^S < \widehat{w}$, and $\rho_R < \rho_I$, as employed above. This establishes that Γ^N is decreasing in r. ■

References

Abernethy, Avery M. and Franke, George R. (1996) “The Information Content of Advertising: A Meta-Analysis.” *Journal of Advertising*, 25(2): 1–17.

Abernethy, Avery M. and Franke, George R. (1998) “FTC Regulatory Activity and the Information Content of Advertising.” *Journal of Public Policy and Marketing*: 17(2), pp. 239–56.

Aaker, David A., Bruzzone DE (1985) “Causes of Irritation in Advertising,” *Journal of Marketing*, 49 (2):47–57.

Almadoss, Wilfred and Sanjay Jain (2015), “Branding Conspicuous Goods: An Analysis of the Effects of Social Influence and Competition,” *Management Science*, 61(9): 2064-2079.

Anderson, Simon and Regis Renault (2006), “Advertising Content,” *American Economic Review*, 96(1): 93-113.

Anderson, Simon and André de Palma (2009), “Information Congestion,” *RAND Journal of Economics*, 40(4): 688-709.

Anderson, Simon and André de Palma (2012), “Competition for Attention in the Information (Overload) Age,” *RAND Journal of Economics*, 43(1): 1-25.

Anderson, Simon and Joshua Gans (2011). “Platform Siphoning: Ad-Avoidance and Media Content,” *American Economic Journal: Microeconomics*, 3: 1-34.

Barry, Thomas E. and Daniel J. Howard (1990), “A Review and Critique of the Hierarchy of Effects in Advertising,” *International Journal of Advertising*, 9(2): 121-135

Becker, Gary and Kevin M. Murphy (1993), “A Simple Theory of Advertising as a Good or Bad,” *Quarterly Journal of Economics*, 108(4): 941-964.

Belanche, Daniel, Carlos Flavián, and Alfredo Pérez-Rueda (2017), “Understanding Interactive Online Advertising: Congruence and Product Involvement in Highly and Lowly Arousing, Skippable Video Ads,” *Journal of Interactive Marketing*, 37: 75-88.

Boik, Andre, Shane Greenstein, and Jeffrey Prince (2019), “The Persistence of Broadband User Behavior: Implications for Universal Service and Competition Policy,” *Telecommunications Policy*, 43(8) 101820.

Bruce, Norris I., Maren Becker, and Werner Reinartz (2020), “Communicating Brands in Television Advertising,” *Journal of Marketing Research*, 57(2): 236-256.

Campbell, Arthur, Dina Maylin, and Jiwoong Shin (2016), “Managing Buzz,” *RAND Journal of Economics*, 48(1): 203-229.

Chandy, Rajesh, Gerry Tellis, Deborah MacInnis, and Pattana Thaivanich (2001), “What to say when: Advertising appeals in evolving markets,” *Journal of Marketing Research*, 38(4): 399–414.

Chen, Yuxin and Qihong Liu (2021), “Signaling Through Advertising When Ad Can Be Blocked,” *Marketing Science*, 41(1):166-187.

Danaher, Peter J. and Tracey S. Dagger (2013), “Comparing the Relative Effectiveness of Advertising Channels: A Case Study of a Multimedia Blitz Campaign,” *Journal of Marketing Research*, L(4): 517-534.

Diamond, Daniel S. (1968), “A Quantitative Approach to Magazine Advertisement Format Selection,” *Journal of Marketing Research*, 5(6): 376-387.

Dukes, Anthony, Qihong Liu, and Jie Shuai (2022), “Skippable Ads: Interactive Advertising on Digital Media Platforms,” *Marketing Science*, 41(3):528-547.

Dukes, Anthony and Esther Gal-Or (2003) “Negotiations and Exclusive Contracts in Advertising," *Marketing Science* 22(2): 222-245.

Du Plessis, Erik (1994), “Recognition Versus Recall,” *Journal of Advertising Research*, 34(3): 75-91.

Fazio, Russell, Paul M. Herr, and Martha C. Powell (1992), “On the Development and Strength of Category-Brand Associations in Memory: The Case of Mystery Ads,” *Journal of Consumer Psychology*, 1(1): 1-13.

Ghosh, Bikram and Axel Stock (2010). “Advertising Effectiveness, Digital Video Recorders, and Product Market Competition,” *Marketing Science*, 29(4): 639-649.

Godes, David, Elie Ofek and Miklos Sarvary (2009). “Content vs. Advertising: The Impact of Competition on Media Firm Strategy,” *Marketing Science*, 28(1): 20-35.

Gritckevich, Aleksandr, Zsolt Katona, and Miklos Sarvary (2018), “Ad Blocking” forthcoming *Management Science*.

Gustafson, P. and S. Siddarth (2007). “Describing the Dynamics of Attention to TV Commercials: A Hierarchical Bayes Analysis of the Time to Zap an Ad,” *Journal of Applied Statistics*, 34(5): 585-609.

Hao, Xinying, Garrett Sonnier, and Frenkel ter Hofstede (2020), “Permission Advertising: Understanding Pre-Roll Ads Leveraging Artificial Empathy,” working paper.

Iyer, Ganesh and Zsolt Katona (2016), “Competing for Attention in Social Communication Markets,” *Management Science*, 62(8): 2304-2320.

Jenkins, Jeffrey, Bonnie Brinton Anderson, Anthony Vance, C. Brock Kirwan, David Eargle (2016), “More Harm Than Good? How Messages That Interrupt Can Make Us Vulnerable,” *Information Systems Research*, Vol. 27(4): 880-896.

Jerath, Kinshuk and Qitian Ren (2021), “Consumer Attention to Favorable and Unfavorable Product Information, and Firm Information Design,” *Journal of Marketing Research*, Vol. 58(2) 343-362.

Joo, Mingyu, Kenneth Wilbur, Bo Cowgill, and Yi Zhu (2014), “Television Advertising and Online Search,” *Management Science*, 60(1): 56–73.

Johnson, Justin (2013), “Targeted Advertising and Advertising Avoidance,” *Rand Journal of Economics*, 44: 128-144.

Ke, Tony, Z.J.M. Shen, and J. Miguel Villas-Boas (2016), “Search for Information on Multiple Products,” *Management Science*, 62(12), 3576-3603.

Keller, Kevin L., and Don R. Lehmann (2009), “Brands and Branding: Research Findings and Future Priorities,” *Marketing Science*, 25(3):740-759.

Kihlstrom, Richard E. and Michael H. Riordan, (1984), “Advertising as a Signal.” *Journal of Political Economy*, 92(3): 427–50.

Lin, Chen, Sriram Venkatraman, and Sandy D. Jap (2013), “Media Multiplexing Behavior: Implications for Targeting and Media Planning,” *Marketing Science*, 32(2), 310-24.

Lin, Song (2022), “The Medium of Advertising,” working paper, Hong Kong University of Science and Technology.

Matejka, F. (2016), “Rationally Inattentive Seller: Sales and Discrete Pricing,” *Review of Economic Studies*, 83: 1125–1155.

Mayzlin, Dina and Jiwoong Shin (2011) “Uninformative Advertising as an Invitation to Search,” *Marketing Science,* 30(4): 666–685.

Milgrom, Paul R. and John M. Roberts (1986), “Price and Advertising Signals of Product Quality.” *Journal of Political Economy*, 94(4): 796–821.

Naik, Prasad and Kalyan Raman (2003), “Understanding the Impact of Synergy in Multimedia Communications,” *Journal of Marketing Research*, 40(4): 375-388.

O’Guinn, Thomas, Chris Allen, and Richard Semenik (2009), *Advertising and Integrated Brand Promotion*, South-Western Cengage Learning; Ohio, United States.

Pashkevich, Max, Sundar Dorai-Raj, Melanie Kellar, and Dan Zigmond (2012), “Empowering Online Advertisements by Empowering Viewers with the Right to Choose: The Relative Effectiveness of Skippable Video Advertisements on YouTube,” *Journal of Advertising Research*, December: 65-71.

Reis, R. (2006), “Inattentive Consumers,” *Journal of Monetary Economics*, 53: 1761–1800.

Resnik, Alan J. and Stern, Bruce L. (1977), “An Analysis of Information Content in Television Advertising,” *Journal of Marketing*, 41(1): 50–53.

Sahni, Navdeep and Harikesh Nair (2020), “Sponsorship Disclosure and Consumer Deception: Experimental Evidence from Native Advertising in Mobile Search,” *Marketing Science*, 39(1): 5-32.

Siddarth, S. and A. Chattopadhyay (1998), "To Zap or Not to Zap: A Study of the Determinants of Channel Switching During Commercials," *Marketing Science*, 17(2):124-138.

Sims, C. (2003), "Implications of rational inattention," *Journal of Monetary Economics*, 50: 665–690.

Singh, Surendra N., Michael L. Rothschild, and Gilbert A. Churchill (1988), "Recognition Versus Recall as Measures of Television Commercial Forgetting," *Journal of Marketing Research*, 25(1): 72-80.

Soberman, David and Yi Xiang (2021), "Designing the Content of Advertising in a Differentiated Market," Forthcoming, *International Journal of Research in Marketing*.

Stern, Bruce L., Dean Krugman, and Resnik, Alan J. (1981), "Magazine Advertising: An Analysis of Its Information Content." *Journal of Advertising Research*, 21(2): 39–44.

Stewart, David W. and Scott Koslow (1989), "Executional Factors and Advertising Effectiveness: A Replication," *Journal of Advertising* 18(3): 21-32.

Teixeira, Thales S., Michel Wedel, and Rik Pieters (2010), "Moment-to-Moment Optimal Branding in TV Commercials: Preventing Avoidance by Pulsing," *Marketing Science*, 29(5): 783-804.

Tellis, Gerard J., Rajesh K. Chandy, and Pattana Thaivanich (2000), "Which Ad Works, When, Where, and How Often? Modeling the Effects of Direct Television Advertising," *Journal of Marketing Research*, XXXVII(1): 32-46,

Todri, Vilma, Anindya Ghose, and Param Vir Singh (2020), "Trade-Offs in Online Advertising: Advertising Effectiveness and Annoyance Dynamics Across the Purchase Funnel," *Information Systems Research*, 31(1): 102-125.

Twedt, Dik Warren (1952), "A multiple factor analysis of advertising readership," *Journal of Applied Psychology*, 36(3): 207-215.

Vakratsas, Demetrios and Tim Ambler (1999), "How Advertising Works: What Do We Really Know?" *Journal of Marketing*, 26(1): 26-43.

Van Zandt T. (2004), "Information Overload in a Network of Targeted Communication," *Rand Journal of Economics*, 35(3): 542-560.

Venkatraman, Vinod, Angelika Dimoka, Paul A. Pavlou, Khoi Vo, William Hampton, Bryan Bollinger, Hal E. Hershfield, Masakazu Ishihara, and Russel S. Winer (2015), "Predicting Advertising Success Beyond Traditional Measures: New Insights from Neurophysiological Methods and Market Response Modeling," *Journal of Marketing Research*, LII(4): 436-452.

Walker, David and Tony M. Dubitsky (1994), "Why Liking Matters," *Journal of Advertising Research*, 34(3): 9-18.

Wedel, Michel and Rik Pieters (2008), "A Review of Eye-Tracking Research in Marketing," *Review of Marketing Research*, 4, Naresh K. Malhotra, ed. Armonk, NY: M.E. Sharpe.

Wu, Yue, Esther Gal-Or, and Tansev Geylani (2022), "Regulating Native Advertising," *Management Science*, forthcoming.

Zhu, Yi and Anthony Dukes (2017), "Prominent Attributes under Limited Attention," *Marketing Science*, 36(5): 683-698.

Tables and Figures

Table 1 Configurations of Signals, Conversions, and Payoffs

Conversion Type	Signal Type & Outcomes (1= Favorable, 0 = Unfavorable)		Payoffs	
	Impression	**Relevance**	**Advertiser**	**Viewer**
Strong	1	1	V_S	u_S
Weak	1	0	V_W	0
None	0	0 / 1	0	0

Table 2 List of Model Parameters

Variable Notation	Description
$I, R \in \{0,1\}$	Informative ad content signals on brand impression and product relevance
$\mathscr{s}$	Either the index for leading signal $\mathscr{s} \in \{I, R\}$ or signal's realization $\mathscr{s} \in \{0,1\}$.
$\rho_{\mathscr{s}} \in [0,1]$	Probability that $\mathscr{s} = 1$, for $\mathscr{s} \in \{I, R\}$
e	Level of investment in entertaining ad content
V_S and V_W	Advertiser's payoff from a strong and weak conversion, respectively
$u_S > u_W = 0$	Viewer's utility from a strong conversion and weak conversion, respectively
$c \in [0, \bar{c}]$	Viewer's attention cost
w, δw, $\widehat{w} = (1-\delta)w$	Viewer's content utility, time-discounted utility, and content utility lost from waiting
$\hat{c}_{\mathscr{s}}^{f}$ and $\tilde{c}_{\mathscr{s}}^{f}$	Viewing cutoffs under format $f \in \{N, S\}$ when signal $\mathscr{s}$ is 1 or 0, respectively
k	Advertiser's investment cost parameter for investment in entertainment

Table 3: Examples of Ad Regimes

Regime	Entertainment Level	Examples
1. Curiosity Inducing ($I \succ R$)	Intermediate	Brand-Focused, New Product Features/Updates
2. Curiosity Inducing ($R \succ I$)	Intermediate	Clickbait, Infomercial
3. Short-Attention Span ($I \succ R$)	Low	Brand Awareness, Reminder
4. Mystery ($R \succ I$)	High	Complex Messaging, Philanthropic and PR

Table A1 Attention Decisions and Conversion Probabilities

Ad Format	First Signal: Type	First Signal: (Un)Favorable Probability	Attention Threshold for $c \in (0, \bar{c})$	Conversion Probability: Strong Conversion	Conversion Probability: Weak Conversion
Non-Skippable	I	ρ_I	$\hat{c}_I^N = \rho_R u_S + e$	$\rho_I \rho_R \hat{c}_I^N / \bar{c}$	$\rho_I[(1-\rho_R)\hat{c}_I^N + (1-\hat{c}_I^N)]/\bar{c}$
		$1-\rho_I$	Irrelevant	0	0
	R	ρ_R	$\hat{c}_R^N = \rho_I u_S + e$	$\rho_I \rho_R \hat{c}_R^N / \bar{c}$	0
		$1-\rho_R$	$\tilde{c}_R^N = e$	0	$\rho_I(1-\rho_R)\tilde{c}_R^N/\bar{c}$
Skippable	I	ρ_I	$\hat{c}_I^S = \rho_R u_S + e - \widehat{w}$	$\rho_I \rho_R \hat{c}_I^S / \bar{c}$	$\rho_I[(1-\rho_I)\hat{c}_I^S + (1-\hat{c}_I^S)]/\bar{c}$
		$1-\rho_I$	All Skip	0	0
	R	ρ_R	$\hat{c}_R^S = \rho_I u_S + e - \widehat{w}$	$\rho_I \rho_R \hat{c}_R^S / \bar{c}$	0
		$1-\rho_R$	All Skip	0	0

Table A2 Optimal Entertainment Levels under Each Format and Signal Order

Ad Format	Optimal Entertainment Level: I	Optimal Entertainment Level: R
Non-Skippable	$e_I^N = \frac{\rho_I \rho_R}{2\bar{c}k}(V_S - V_W)$	$e_R^N = \frac{\rho_I}{2\bar{c}k}[\rho_R V_S + (1-\rho_R)V_W]$
Skippable	$e_I^S = \frac{\rho_I}{2\bar{c}k}[\rho_R(V_S - V_W) - r]$	$e_R^S = \frac{\rho_R}{2\bar{c}k}(\rho_I V_S - r)$

A: Magazine Ad

B: Clickbait Ad

Figure 1: Illustration of the Types of Ad Content

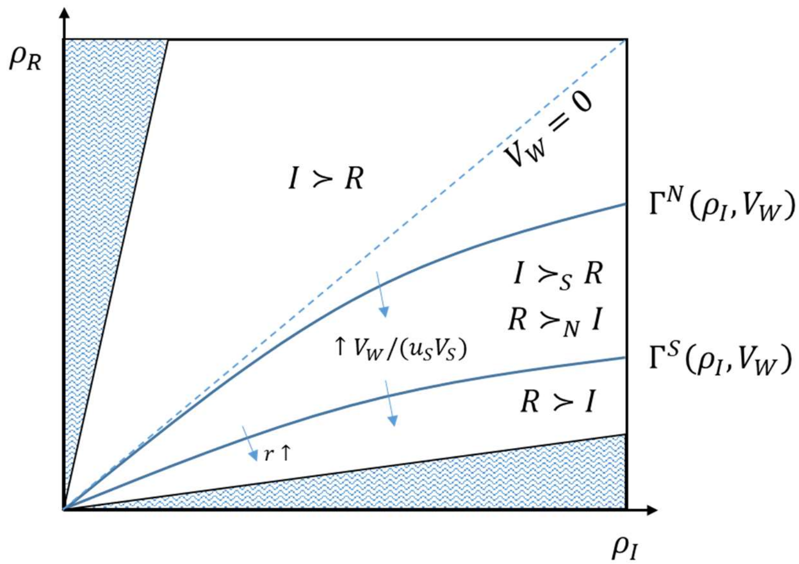

Figure 2 Optimal Informational Content Strategy

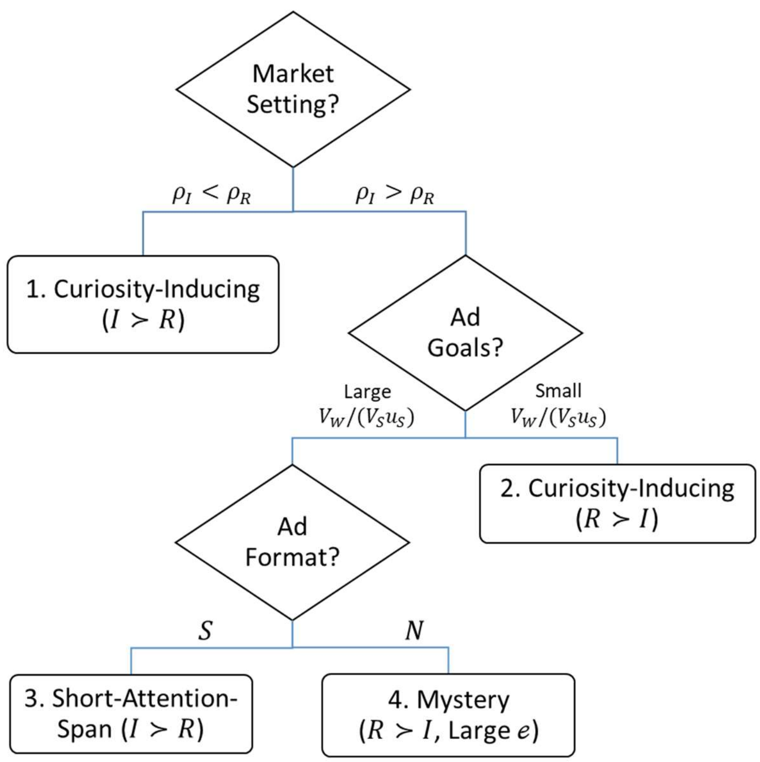

Figure 3 Guiding Motives for Optimal Ad Content Strategies

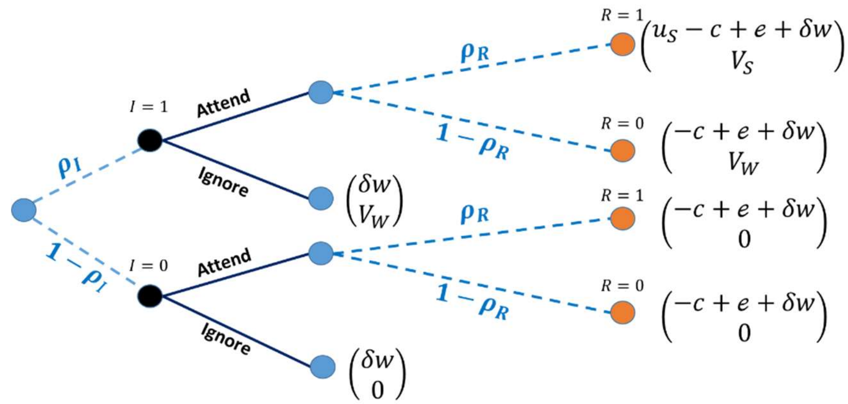

Figure A1: Decision Tree with Payoffs $\begin{pmatrix} \text{Viewer} \\ \text{Advertiser} \end{pmatrix}$ when I is First Signal

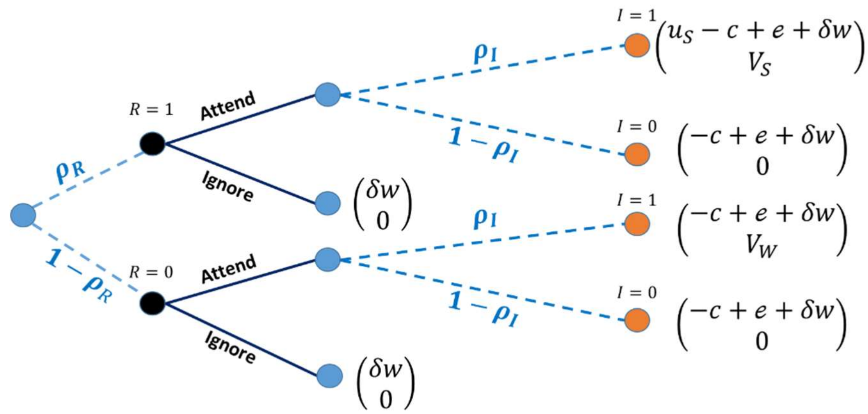

Figure A2: Decision Tree with Payoffs $\begin{pmatrix} \text{Viewer} \\ \text{Advertiser} \end{pmatrix}$ when R is First Signal