Article

AMERICAN MARKETING ASSOCIATION

Variety Effects in Mobile Advertising

Journal of Marketing Research
1-21

DOI: 10.1177/00222437211056090
journals.sagepub.com/home/mrj
SAGE

Omid Rafieian and Hema Yoganarasimhan

Abstract

Mobile app users are often exposed to a sequence of short-lived marketing interventions (e.g., ads) within each usage session. This study examines how an increase in the variety of ads shown in a session affects a user's response to the next ad. The authors leverage the quasi-experimental variation in ad assignment in their data and propose an empirical framework that accounts for different types of confounding to isolate the effects of a unit increase in variety. Across a series of models, the authors consistently show that an increase in ad variety in a session results in a higher response rate to the next ad: holding all else fixed, a unit increase in variety of the prior sequence of ads can increase the click-through rate on the next ad by approximately 13%. The authors then explore the underlying mechanism and document empirical evidence for an attention-based account. The article offers important managerial implications by identifying a source of interdependence across ad exposures that is often ignored in the design of advertising auctions. Furthermore, the attention-based mechanism suggests that platforms can incorporate real-time attention measures to help advertisers with targeting dynamics.

Keywords

mobile advertising, variety, diversity, attention, digitization, sequential decision making

Online supplement: https://doi.org/10.1177/00222437211056090

The smartphone industry has seen unprecedented growth over the past decade, with more than three billion worldwide users in 2020 (eMarketer 2020a). In 2019, an average U.S. adult spent almost four hours per day on mobile devices, surpassing the time spent on television for the first time (eMarketer 2019). This growth in smartphone adoption and usage has made the mobile medium attractive for marketing interventions. These interventions are often short-lived (e.g., mobile ads, multimedia messaging service, push notifications) and provide marketers with many opportunities to interact with users. As such, mobile users are exposed to a variety of marketing interventions within a short period of time.

As the variety of marketing interventions increases, managers and researchers need to understand the consequences of this increased variety in the mobile ecosystem. To date, the literature has viewed the increased variety of interventions as a means to explore (vs. exploit) and learn about consumers' tastes (Lattimore and Szepesvári 2020), increase fairness (Dwork et al. 2012), prevent polarization (Celis et al. 2019), and increase reachability (Dean, Rich, and Recht 2020). However, it is not clear how the increased variety itself affects consumer behavior. Understanding the behavioral consequences of increased variety is particularly relevant for managers and researchers because it has important market design implications for the platforms that design marketplaces for consumers and marketers to interact.

We study the effects of increased ad variety on consumer behavior in the context of mobile in-app advertising, the most popular type of marketing intervention on mobile devices. It is now the dominant advertising channel in the United States and generates more than $100 billion in ad spending (eMarketer 2020b). A common feature of in-app advertising is the use of refreshable ad slots: each ad exposure lasts a short time (e.g., one minute), and the slot is then refreshed with another ad. Thus, even in the course of a short ten-minute session, a user can be exposed to a variety of ads. This feature of in-app advertising makes it particularly suitable for studying how variety of a sequence of ads seen earlier in a session affects a user's responsiveness to the current ad. Specifically, we aim to answer the following questions:

1. How does a unit increase in ad variety in a session influence users' responsiveness to the next ad?
2. What is the underlying mechanism that explains these effects (if any)?

Omid Rafieian is Assistant Professor of Marketing and Demir Sabanci Faculty Fellow of Marketing and Management, SC Johnson College of Business and Cornell Tech, Cornell University, USA (email: or83@cornell.edu). Hema Yoganarasimhan is Professor of Marketing and Michael G. Foster Faculty Fellow, University of Washington, USA (email: hemay@uw.edu).

3. What are the managerial implications of these findings for platforms?

To answer these questions, we use a large-scale data set from the leading mobile in-app ad network from a large Asian country. Two key features of the ad network make it well suited to study variety effects. First, like most in-app ad networks, it employs a refreshable ad format; each ad lasts one minute and is followed by another ad exposure. Second, it uses a probabilistic auction to allocate ads, which allows for a wide variety of ads to be shown in a single session, which, in turn, gives us unconfoundedness in ad assignment. Note that unconfoundedness is a necessary condition for causal inference and is often missing in observational studies on advertising. These two features provide the necessary requirements to draw causal inference and answer our research questions.

We face three key challenges in answering our research questions. First, our treatment variable—an increase in the ad variety of a session—is not fully randomized. Some impressions have a higher propensity to be assigned to the treatment than others due to pretreatment characteristics, such as targeting variables and the set of prior ads. We refer to this as the "pretreatment confounding problem." Second, we are interested in the outcome of the next ad exposure after the treatment. However, some users may leave the session right after receiving the treatment and before we observe the outcome. If their decision to leave is a function of the treatment, this censoring can interfere with inference. We refer to this as the "dynamic selection problem." Third, during the posttreatment phase—from treatment assignment to the outcome collection phase—other variables can also covary with our treatment (e.g., the identity of the ads shown, their recency and frequency), which makes isolating the treatment effect of an increase in ad variety challenging. We call this issue "posttreatment confounding" because our treatment definition is not separate from some other posttreatment factors that might also affect users' decisions.

We present a methodological framework that uses the unconfoundedness of ad assignment and helps us address the aforementioned challenges. First, to account for nonrandomness in assignment to an increase in ad variety (pretreatment confounding), we show that the unconfoundedness of ad assignment implies the unconfoundedness of our treatment. That is, conditional on observables, treatment assignment is as good as exogenous. Specifically, we employ the main insight from Rosenbaum and Rubin (1983) and estimate the propensity score for being assigned to an increase in ad variety for any given impression. We then assess covariate balance for these propensity scores and feed them into the main regression model to control for pretreatment confounding.

Second, to address our dynamic selection challenge, we employ an imputation strategy in which we impute the observation in the next period for users who left the session after being assigned to a treatment (Little and Yau 1996). Under the unconfoundedness of ad assignment, we can accurately impute the ad that would have been shown had the user not left the session because we can estimate the distribution of ad assignment given observables. In particular, we use a specific feature of our setting, which ensures that auctions for two impressions that happen around the same time and share the same targeting characteristics are identical. This allows us to use a complementary sample of auctions that occur around the same time as our missing impressions and impute the missing ad assignments.

Our solution to the third challenge of posttreatment confounding is based on simple logic: we want to ensure that our estimates of treatment effect only capture our treatment, namely, an increase in ad variety of the session. As such, we need to control for any posttreatment information that imperfectly covaries with our treatment. Because our posttreatment phase spans two consecutive exposures (often referred to as exposure $t-1$ and t), where the treatment is assigned in the former and the outcome is collected in the latter, we control for the fixed effects of the specific ads shown in these two exposures, as well as the frequency and recency of the ad shown in the later exposure. This allows us to achieve a ceteris paribus interpretation under some mild conditions. That is, if our treatment effect estimate is β, the interpretation would be as follows: a unit increase in the ad variety of a session increases the outcome of the next ad by β, holding all else constant.

We use an inverse probability weight-adjusted regression to estimate our main effects. We find that an exogenous increase in the ad variety of a session results in a significantly higher click-through rate (CTR) on the next ad, holding all else fixed. The magnitude of our treatment effect accounts for approximately 13% of the average CTR, which implies that variety effects are relatively sizable. We then consider a series of alternative specifications to examine the robustness of our findings. Notably, we consider a restrictive exact-matching model, in which we match the exact sequence of ads shown in a session except for the ad shown in the treatment stage (e.g., matching from treatment to from the control) and then control for the fixed effects of the ad that is different to fully isolate the effect of an increase in ad variety. Across all the robustness checks, our results show the same pattern: increasing ad variety generates more clicks on the next exposure.

Next, we explore the mechanism underlying our main findings. In our analysis, we focus on the main feature that differs across our treatment and control groups, namely, the novelty of the ad shown in the treatment phase. The behavioral literature has demonstrated that showing more novel stimuli in a given space increases users' attention to that space (Helson 1948; Kahneman 1973). Therefore, we develop an attention-based explanation for variety effects, wherein the novelty of prior ads increases users' attention to the advertising slot, which in turn increases their likelihood of clicking on the next ad.

We conduct a series of empirical tests to examine the validity of the predictions resulting from this mechanism. First, we use the fact that the novelty of the control condition can vary on the basis of the within-session frequency and recency of the ad shown in the treatment phase: the more frequent and recent the ad is, the less novel we expect it to be. We show that as the novelty of the control group drops, the treatment effects become larger. Second, users with higher levels of presession exposure to ads and those with more recent presession ad exposures exhibit smaller treatment effects. This is consistent with the idea that the novelty of within-session interventions deteriorates with higher and more recent ad exposures before the session. Third, if our attention-based account is correct, the treatment effect should be smaller when past variety is high because the novelty of an increase in variety is less likely to have an impact on user attention in this case. We test this prediction and demonstrate that the treatment effects are indeed smaller when past variety is already high, and vice versa. Together, these tests provide empirical support for the validity of the proposed mechanism.

In summary, our study contributes to the literature in several ways. Substantively, the main contribution lies in establishing the causal effects of variety in the advertising context. To our knowledge, ours is the first research to study the downstream effects of an increase in ad variety on consumers' responses.[1] The findings broaden our understanding of how constructs such as variety and diversity affect consumer behavior, which is of critical importance to digital platforms as they employ increasingly experimental techniques and commit to increasing ad diversity. Furthermore, we propose an attention-based mechanism to explain the effects of increased ad variety, an important finding because it provides a parsimonious and testable account of the effects of variety and highlights the importance of attention-based measures. A key implication for platforms is to incorporate these attention-based measures in their targeting offerings to advertisers and in their quality-scoring system to improve the performance of their auctions. From a methodological standpoint, our article develops an empirical framework to study the effects of variety in sequential settings that can be applied to other domains (e.g., music-streaming services). Our empirical framework is fairly general and requires only the unconfoundedness assumption, which can be easily satisfied by a digital platform. We expect our substantive findings and empirical framework to be relevant for digital platforms within and outside the advertising domain.

Related Literature

First, our study relates to the marketing literature on variety. The concept of variety has been examined through two broad viewpoints. The first stream views variety (in consumption) as an outcome variable and studies consumers' variety-seeking behavior (McAlister 1982; Ratner, Kahn, and Kahneman 1999) and, more broadly, their demand for product variety (Datta, Knox, and Bronnenberg 2017; Kim, Allenby, and Rossi 2002). In the second stream, variety serves as a factor influencing some outcome variables associated with consumer behavior, such as the link between assortment variety and store choice (Hoch, Bradlow, and Wansink 1999), content variety and consumers' engagement (Redden 2008), and the word-of-mouth dispersion and television ratings (Godes and Mayzlin 2004). In line with the second stream, our article studies the effects of increasing ad variety on consumers' ad response. Our work contributes to this literature in two ways. From a substantive perspective, we establish variety effects in the context of advertising and propose an attention-based account to explain these effects that can be applied in other domains. From a methodological perspective, we provide an empirical framework to study the effects of variety in sequential treatment settings.

Second, our research relates to the literature on advertising marketplaces that adopts a platform perspective to study questions related to advertising. Work in this domain often focuses on broad questions of market design (Choi and Mela 2019; Yao and Mela 2011), platforms' incentives to provide tools such as granular behavioral targeting (Rafieian and Yoganarasimhan 2021) or ad avoidance technology (Tuchman, Nair, and Gardete 2018; Wilbur 2008), and advertising externalities that affect market design (Wilbur, Xu, and Kempe 2013). Our article extends this literature by recognizing a new type of externality created by ad variety, which is of relevance to advertising marketplaces that seek to incorporate diversity and fairness criteria in their decision making. In addition, we highlight important challenges in designing auctions and optimal adaptive experimentation systems in light of this externality and offer potential solutions.

Finally, our article relates to the literature on advertising dynamics. Prior literature on television advertising has extensively documented evidence for different dynamic effects, such as ad avoidance, carryover effects, wear-in, wear-out, and an S-shaped ad response curve (Aravindakshan, Peters, and Naik 2012; Danaher 1995; Naik, Mantrala, and Sawyer 1998; Tellis 2003). While this body of work focuses on aggregate response models, a series of recent studies on digital advertising take advantage of individual-level data and document temporal effects—such as carryover effects and wear-out (Chae, Bruno, and Feinberg 2019), the effect of multiple ad creatives (Braun and Moe 2013; Bruce, Murthi, and Rao 2017), advertising avoidance (Deng and Mela 2018; Wilbur 2016), spillover effects (Jeziorski and Segal 2015; Rutz and Bucklin 2011; Sahni 2016), and effects of temporal spacing (Sahni 2015)—and, more broadly, attribution dynamics (Danaher and Van Heerde 2018; Li and Kannan 2014; Zantedeschi, Feit, and Bradlow 2017). For a summary of individual-level models of digital advertising, see Bucklin and Hoban (2017). Our article extends this stream of literature in two ways. First, we establish a new dynamic effect, namely, the effect of ad variety on user behavior. Second, we offer a new behavioral account of consumers' ad response based on attention and adaptation-level theory that can be used more broadly in advertising studies. Finally, we provide a methodological framework

[1] The only prior study related to variety in this context is Schumann et al.'s (1990), in which they show that variation of ad content over a repeated advertising schedule increases users' responsiveness to that ad. While the authors focus only on variation in the content for one ad in a lab setting, we extend this to a variety of potentially competing ads in a large-scale in-app advertising market.

that can be used in other studies on advertising dynamics under some mild assumptions.

Setting and Data

Setting

Our data come from the leading mobile in-app ad network form a large Asian country, which had over 85% market share at the time of this study. The ad network functions as a matchmaker between advertisers and mobile apps and serves ads inside mobile apps. The scope and scale of the network are large, and it generates more than 50 million ad impressions daily. We begin by describing the four main stakeholders in this marketplace:

- *Users*: Users are the individuals who generate impressions through use of their mobile apps. For each ad impression shown, the user can decide whether to click on the ad or not.
- *Publishers*: Publishers are the app developers who have joined the ad network. Their revenue comes from clicks on the ads shown in their apps. This is the main monetization strategy for most of the apps in our data.
- *Advertisers*: Advertisers are the firms (usually mobile apps) that want to show their ads to mobile app consumers. Advertisers create banner ads (in either JPEG or GIF formats) and submit a bid that indicates both their willingness to pay per click. Advertisers can also target ads based on the following variables associated with an impression: province, app category, hour of the day, smartphone brand, mobile service provider (MSP), and connectivity type.
- *Ad network*: The ad network is the platform that facilitates the matching between ads and impressions generated by a user–app combination. It runs a real-time auction to select the ad to show in each impression. In our context, the platform has full control over ad format (only banner ads in the bottom of an app's screen) and allocation through the auction. It operates based on a cost-per-click (CPC) revenue model, which means that it generates revenue only when a click happens. The ad network has strong economic incentives to understand the drivers of clicks and to employ this understanding to generate more revenue.

The ad network in this study employs refreshable ad slots, where each impression lasts one minute.[2] When a user starts using an app, the ad network runs an auction to determine the winning ad and serves this ad for one minute. If the user continues using the app beyond one minute, the ad network treats this as a new impression and runs another auction to determine the next ad to show the user. The practice of using a refreshable ad slot helps create more ad impression opportunities by reducing the time allocated to each impression. Figure 1 presents a visual representation of the ad format, the ad slot, and the sequential nature of ad delivery in our setting.

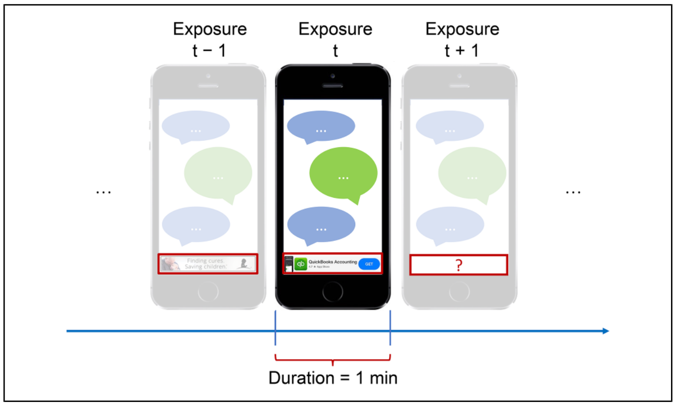

Figure 1. A visual representation of the ad slot in our setting.
Notes: The highlighted rectangle at the bottom of app is the refreshable ad slot.

Data

We have data on all the impressions and corresponding clicks (if any) in the platform for all participating apps for a one-month period (September 30, 2015–October 30, 2015). For each impression, we observe the following information: (1) time and date, (2) AAID (user identifier), (3) app ID (publisher), (4) ad ID, (5) bid submitted by the winning ad, (5) GPS information (including the exact latitude and longitude of the user), (6) click indicator, and (7) targeting variables that contain the province, app category, hour of the day, smartphone brand, connectivity type, and MSP. Notably, all the variables that the advertiser can target are observable to us.[3] Thus, we are able to avoid many of the common problems related to endogeneity in the measurement of ad effects due to targeting based on unobservables, as we discuss in greater detail subsequently.

Overall, the scale of our data is large, with more than 1,594,831,699 impressions over a one-month period and an average of 600 auctions per second. Next, we explain how we sample from the data and describe an important aspect of the data-generating process, namely, the auction mechanism.

Session Definition and Sampling

Recall that our goal is to study the effects of an increase in ad variety in a session on a user's response to the current ad. As such, the definition of a "session" is central to our study. Because we focus on an increase in ad variety in a session, we want the user in a session to be able to recognize this

[2] It is important for the length of all the exposures to be the same because we know that the length of ad exposures affects ad responsiveness in digital advertising settings (Danaher and Mullarkey 2003).

[3] The ad network also has access to the internet service provider (ISP) for each impression if it happens over Wi-Fi. However, in our data, this information is missing for the majority impressions, and advertisers do not seem to use it for targeting. Thus, we do not use the ISP in our analysis.

increase. Thus, we need to avoid long usage gaps in our session definition to ensure that the user is able to recognize an increase in ad variety. To that end, we define a session as a set of consecutive impressions generated by a user within an app, such that the gap between two consecutive impressions is less than 10 minutes.[4]

Given the substantive goal of our research, we construct a sample of users for whom we have the entire behavioral history on the platform. This excludes the sample of users whose past activity comes before the beginning of our data and those whose activity logs are not stored in the platform server at different points. We then focus on the top app and collect data from all sessions with our sample of users using this app.[5] Overall, our sample comprises 85,450 users who generated 1,197,850 sessions and 6,805,322 impressions in the messenger app. We observe a total of 327 unique ads in this sample. The length of the sessions in our sample varies quite a bit. While half of the sessions end after the first two exposures, over 25% last five or more exposures (see Figure W1 in Web Appendix A). All the descriptives in this section are shown for this sample. However, we use the data from other users and apps to supplement our analysis. For the sample of users we focus on, we keep track of the data generated by these users in other apps to segment them on the basis of their behavioral history and explore the heterogeneity in their responsiveness to ad variety. Furthermore, we use the impressions from other users who are not in our sample in the top app for imputation purposes in our estimation procedure. We discuss these uses in greater detail subsequently.

Finally, although we use the same data source as Rafieian and Yoganarasimhan (2021), the specific sample and the goal of the studies differ considerably. Rafieian and Yoganarasimhan use a random sample of users over a three-day span (October 28–October 30) to predict CTRs across all apps and ads. They use the earlier data to construct detailed behavioral and contextual features and then examine the platform's incentives to allow more granular targeting in counterfactual scenarios. Thus, they focus on the interplay between the platform's market design and advertisers' bidding decisions in a two-sided market. We also have an entirely different goal in this article: specifically, we want to examine a consumer-level effect when the platform increases ad variety. Furthermore, we use a completely different sample for which we can make sure we have the entire behavioral history for all our users and focus our analysis on the data from the top app.

Data-Generating Process

We next describe the data-generating process in our setting. Let i denote a session, and let t denote an exposure number within a session. Each exposure t in session i comes with three pieces of information: (1) impression-level characteristics ($X_{i,t}$) that capture all the observable attributes associated with the user and context of the impression (e.g., the smartphone brand), (2) the ad shown in the exposure ($A_{i,t}$), and (3) the user's decision to click on the ad shown in the exposure ($Y_{i,t}$).

The ad network uses a quasi-proportional auction to allocate ads to impressions (Mirrokni, Muthukrishnan, and Nadav 2010). The main distinction between a quasi-proportional auction and other commonly used auctions (e.g., second price or Vickrey) is the use of a probabilistic winning rule—that is, all ads participating in the auction for an impression have a nonzero probability of winning. For exposure t in session i, the probability that ad a wins this impression is given by

$$\pi_{i,t}(a) = 1(a \in \mathcal{C}_{i,t}) \frac{b_{i,a} m_{i,a}}{\sum_{k \in \mathcal{C}_{i,t}} b_{i,k} m_{i,k}}, \tag{1}$$

where $\mathcal{C}_{i,t}$ is the set of ads participating in the auction for this exposure and $b_{i,a}$ and $m_{i,a}$ are ad a's bid and quality score in session i, respectively. Thus, the variation in $\pi_{i,t}(a)$ can stem from variation in $\mathcal{C}_{i,t}$, $b_{i,a}$, $m_{i,a}$, or some combination of these variables. The variation in $\mathcal{C}_{i,t}$ is largely driven by advertisers' targeting decisions. For example, if ad a chose not to target the province where the user in session i is located, then ad a will not belong to $\mathcal{C}_{i,t}$, which implies $\pi_{i,t}(a) = 0$. The quality score $m_{i,a}$ is a measure of profitability that the platform assigns to ad a in session i. The extent of customization in the quality scores is low: the ad network simply assigns one aggregate quality score to each ad and only updates it once a day. While $m_{i,a}$ can vary across sessions, the extent of variation is low. Finally, $b_{i,a}$ is ad a's bid for session i. In our setting, each ad could submit only one bid at a given time of the day.[6]

In summary, $\mathcal{C}_{i,t}$, $b_{i,a}$, and $m_{i,a}$ together determine the distribution of propensity scores ($\pi_{i,t}(a)$) for ads for each exposure in session i. As such, the ad shown at exposure t, $A_{i,t}$, is a draw from this probability distribution. This probabilistic allocation rule thus gives us random variation in ad assignment across and within sessions, which form a core part of our identification strategy. It is worth emphasizing that the extent of exogenous variation created by the auction only helps us without limiting

[4] We do not assume that a click automatically ends a session for two reasons. First, empirically, we observe that 80% of users who click come back to the app within ten minutes. Second, there is no theoretical reason to believe that a click will affect the user's memory of prior ads if that user comes back to the app within a short time. That said, our results are robust to alternative definitions of a session—for example, when we assume a click ends a session, or when we allow for larger or smaller gaps between consecutive impressions.

[5] Our choice of using the top app is only for cleaner analysis, so that we can better control the context and app-switching. This allows us to only sample sessions that are entirely in one app. The main results are robust when we consider other apps.

[6] Advertisers could not customize bids by targeting variables. For example, they could not submit different bids for two provinces at the same time, even if their willingness to pay for the clicks in the two provinces was different. Furthermore, if an advertiser changes its bid at some point in time, it is updated for all the sessions that start in the next hour of the day.

Table 1. Summary Statistics of the Targeting Variables.

Variable	Number of Subcategories	Share of Top Subcategories			Total Number of Impressions
		First	Second	Third	
Province	31	24.67%	9.61%	7.45%	6,805,322
Hour of the day	24	8.43%	7.98%	7.21%	6,805,322
Smartphone brand	7	44.62%	38.18%	8.62%	6,177,053
Connectivity type	2	50.33%	49.67%		6,805,322
MSP	3	50.18%	43.98%	5.84%	6,635,836

Notes: The last column shows the total number of nonmissing observations for each variable. While information about province and hour of the day is always available, other variables are missing for some impressions. We computed the shares shown after excluding the missing observations for each variable.

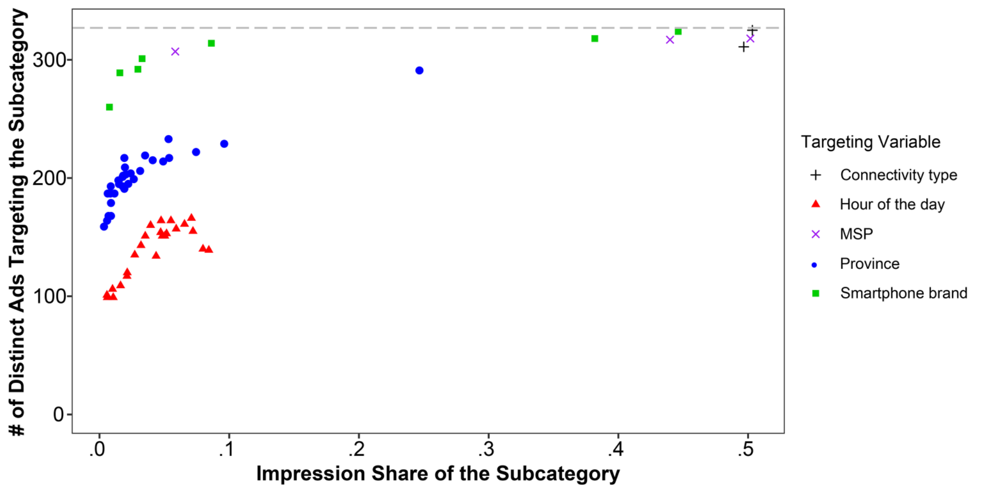

Figure 2. Relationship between the number of distinct ads targeting a subcategory and the impression share of that subcategory within the targeting category.
Notes: All subcategories within each targeting category are in the same color and shape. The dashed gray line on the top is the total number of distinct ads available in our data (which is 327).

our ability to extend our results to other more commonly used auction settings.[7]

Summary Statistics

Targeting Variables

Targeting variables are the dimensions on which advertisers can target their ads. In our setting, these consist of province, app category, hour of the day, smartphone brand, MSP, and connectivity type. All these variables are categorical, and advertisers can specify which subcategories they want to show their ads in (e.g., an advertiser can indicate that it wants its ads to be shown only from 6 P.M. to 9 P.M. every day on Samsung phones in one specific province). We first report the impression share of the top three subcategories within each targeting variable in Table 1.

Next, we examine the extent of targeting that occurs in the platform. We know that if a subcategory (e.g., Samsung in the category of smartphone brand) is excluded by an ad through its targeting decision in a given hour of the day, then zero impressions of that ad will be shown in that subcategory during that period. Thus, the number of distinct ads shown within a subcategory is informative of the number of advertisers targeting that subcategory. Furthermore, we can correlate the impression share of a targeting subcategory with the number of distinct ads targeting it to illustrate the relationship between a subcategory's popularity and the extent to which it is targeted. Therefore, for each subcategory within a targeting variable/dimension, we first calculate the number of distinct ads that show at least one impression in that subcategory. Then, we plot the number of distinct ads targeting it against its share of impressions in that subcategory. The results of this analysis appear in Figure 2.

Three important patterns emerge in Figure 2. First, we find that some variables are not used much for targeting. In particular, all the subcategories within connectivity type and MSP are close to the gray dashed line at the top. This implies that most advertisers were showing their ads irrespective of these

[7] It is similar to the case in which we want to generalize the results from an experiment in Facebook to the setting with their auctions in place. For a description of commonly used auction mechanisms in digital advertising, see Tunuguntla and Hoban (2021).

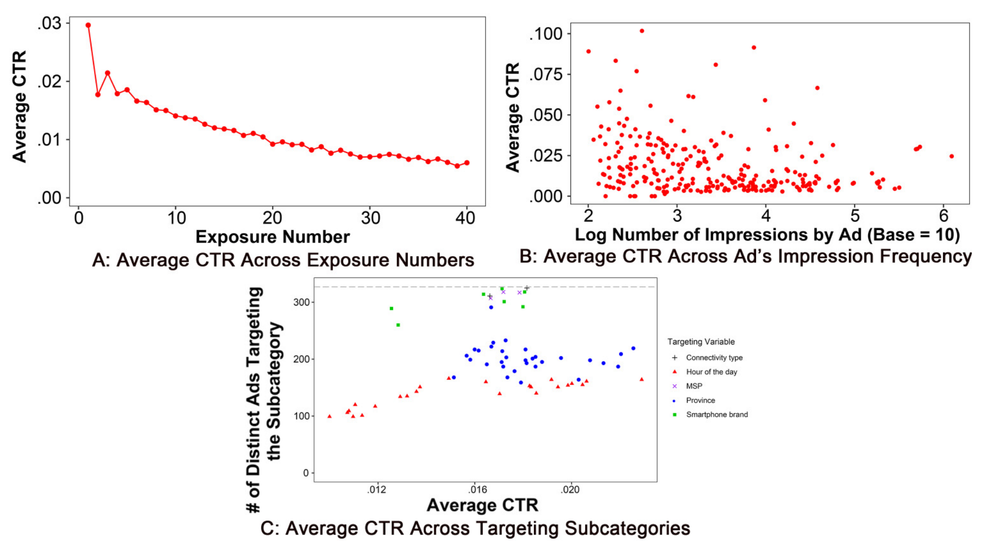

Figure 3. Variation in CTR.
Notes: Measure of CTR is in absolute terms, not percentage.

variables. While the subcategories in the smartphone brand differ slightly in the number of ads targeting them, the extent of targeting is still limited. We find that province and hour of the day are the main variables used for targeting: all subcategories within these variables are considerably different in terms of the number of distinct ads targeting them. The second insight from Figure 2 also relates to this difference: subcategories with a higher share of impression within a category seem to be more popular among advertisers. For example, the circle in the top center denotes the largest province in the country (and contains the capital city) with the highest share of impressions. This province also has the highest number of distinct ads targeting it (compared with other provinces). In contrast, the triangles in the bottom left are midnight hours that have the lowest impression shares and fewer advertisers targeting them. However, even these unpopular hours attract a lot of ads (more than 100). This brings us to the third key insight from Figure 2: there is no niche targeting in this market. We can expect a significant amount of within-session variation in the set of ads in most sessions. This, in turn, facilitates the study of variety effects on user behavior.

Variation in CTR

Clicks are the main outcome of interest in this study and are important in our setting for a few key reasons. First, we view the problem through the lens of a platform that runs a CPC auction and wants to generate more clicks. Thus, clicks are directly tied to platform revenues. Second, the majority of ads in our setting are mobile apps interested in app-installs. These ads are often referred to as "performance ads" because they have objective performance measures, as opposed to "brand ads" whose goal is often to generate more reach and brand recognition (Arnosti, Beck, and Milgrom 2016). While a click on a brand ad is often not informative regarding the final conversion outcome (e.g., click on a Ford ad takes the user to Ford's main home page), a click on a performance ad is a direct step toward conversion. For example, in our context, a click on the advertised app takes the user to the app store page, which is one click away from conversion. Thus, a click serves as a strong engagement signal of the ultimate outcome for ads in our study.

Figure 3 presents some descriptives on the variation in CTR across (1) exposure numbers, (2) ads, and (3) sessions. We start with the variation in CTR across exposure numbers, which is the most basic graph that illustrates the within-session variation in CTR. Figure 3, Panel A, reveals a downward trend in average CTR across exposure numbers. This suggests that it is essential to control for exposure number when studying the effects of within-session interventions such as variety.[8] Next, we focus on the variation in average CTR across ads. In Figure 3, Panel B, we plot each ad's average CTR across the log of the number of impressions of the ad. This allows us to visualize the dispersion in ad-specific CTR at different levels of impression frequency. Overall, we observe considerable variation in ad-specific CTR at all frequency levels, especially across low-frequency ads. Finally, we examine variation in CTR across sessions with respect to their targeting variables. As such, Figure 3, Panel C, plots the number of distinct ads targeting each subcategory across their average CTR. This is the equivalent of Figure 2, the difference being that the x-axis is the average CTR of each subcategory as opposed to impression share. First, considering only the x-axis, we find substantial variation in CTR across provinces and hours of the day. However, variation in CTR is minimal in the subcategories of other

[8] The average CTR across all the impressions (for all the exposure numbers) is .0174.

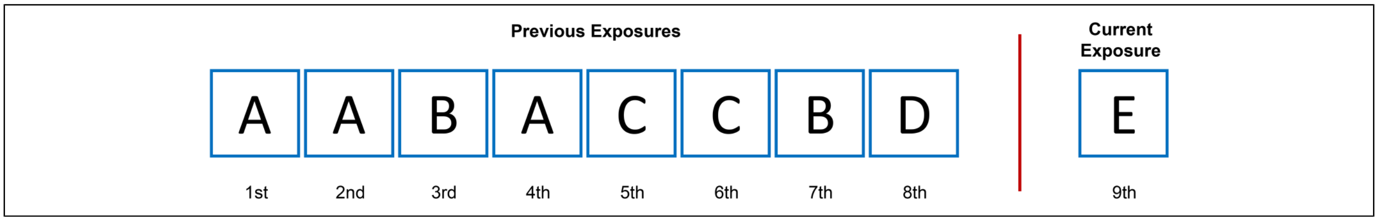

Figure 4. An example of a session in which the user is at the ninth exposure.
Notes: The numbers represent exposure number t, and each rectangle represents the ad shown in that exposure. The letter coding refers to ad IDs; each letter represents one unique ad. For example, the user is shown the same ad (coded in letter A) during the first, second, and fourth exposures.

targeting variables, such as MSP and connectivity type. In general, the variation in CTR shrinks as y-axis values increase, where almost all ads target all subcategories. Intuitively, this makes sense because advertisers want to use variables for targeting that are informative about CTR. Finally, we observe a small positive correlation between the number of distinct ads targeting a subcategory and the corresponding CTR, which suggests that subcategories with higher CTRs are more popular among advertisers. This pattern is indicative of a selection problem, which we discuss in greater detail subsequently.

Preliminary Analysis

A Simple Measure of Variety

Quantifying variety is challenging because consumers' perceptions of variety vary depending on the context and the information structure of the assortment (Hoch, Bradlow, and Wansink 1999). Different measures in the literature capture certain aspects of variety in a set of objects, such as breadth of variety, diversity, or concentration. We want to measure variety for the sequence of prior ads shown in a session. Figure 4 presents an example of such a sequence, where a user has seen a sequence of eight ads and is now at the ninth exposure.

In this section, we focus on the simplest conceptualization of variety over a sequence of ads—namely, breadth of variety, which counts the number of distinct ads shown. In the example in Figure 4, the breadth of variety is four, as there are four different ads shown in the sequence of eight prior exposures. We define the sequence of ads shown in session i as $\langle A_{i,t}\rangle_{t=1}^{T_i}$, where $A_{i,t}$ is the ad shown in the t^{th} exposure in session i and T_i is the total number of exposures shown in session i. We define the breadth of variety as follows:

$$V_{i,t} = |\{A_{i,1}, \ldots, A_{i,t-1}\}|. \quad (2)$$

In addition to simplicity, another advantage of this measure is that we can decompose it into binary pieces. We subsequently use this feature of $V_{i,t}$ to define our empirical problem.

Variation in Breadth of Variety

We now present descriptive statistics on the distribution of breadth of variety in our data. Conceptually, variation in breadth of variety in our setting stems from the probabilistic allocation mechanism. Figure 5, Panel A, presents the empirical distribution of breadth of variety for the ninth impression in the sessions in our data. We observe considerable variation in users' exposure to variety: all levels of variety, from one to eight, occur in the data. This is promising because we need sufficient variation in variety to conduct a meaningful analysis.

While we have sufficient variation in users' exposure to variety at any given exposure,[9] it is not clear whether this variation is distributed identically across different sessions. Indeed, we know that targeting subcategories vary significantly in their popularity based on advertisers' targeting decisions (see Figure 2 and Figure 3, Panel C). As such, we expect sessions with more appealing targeting characteristics to have a higher variety of ads simply because they have more ads in their inventory. We now use this intuition to examine differences in the distribution of variety across sessions as a function of their targeting popularity. We define the targeting popularity of an impression as the number of ads that are targeting all the subcategories associated with that impression. We then use a median split to divide impressions at the ninth exposure into two subsets: popular and unpopular impressions. Figure 5, Panel B, shows the distribution of variety for each subset. This figure confirms our intuition that the two distributions are different: more popular impressions show a higher variety of ads with a gap of more than one point in the means of these distributions. Thus, in our main analysis, we need to ensure that we separate the effects of variety from targeting popularity.

Preliminary Results

We next run a series of regressions to explore the relationship between the click outcome and the variety of previous ads. Let $Y_{i,t}$ denote the click outcome for session i at exposure t, let $X_{i,t}$ denote the set of presession variables (e.g., province, hour of the day),[10] and let $H_{i,t}$ be the sequence of all ads shown in the session (i.e., $H_{i,t} = \langle A_{i,1}, \ldots, A_{i,t}\rangle$). For our preliminary

[9] We only show the histogram for the ninth exposure, but coverage is the same for other exposures.
[10] The reason we have subscript t for presession variables is that some variables, such as hour of the day, can actually change within the session.

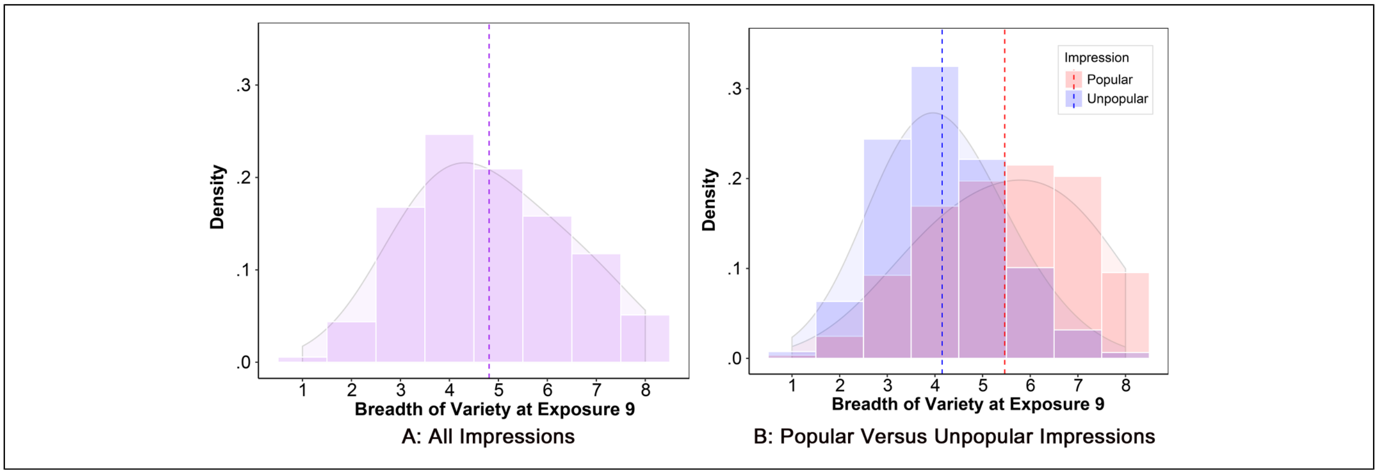

Figure 5. Distribution of breadth of variety at the ninth exposure for all impressions and subsets of popular and unpopular impressions.
Notes: Popularity is defined as the number of ads that targeted an impression based on the targeting characteristics available.

analysis, we estimate the following regression model:

$$Y_{i,t} = \beta V_{i,t} + f(X_{i,t}, H_{i,t}) + \epsilon_{i,t}, \tag{3}$$

where β captures the marginal effect of the breadth of variety of previous ads on the user's click probability and $f(X_{i,t}, H_{i,t})$ can be any nonparametric function that separates the effects of the other covariates on the outcome from the effects of variety. In our preliminary analysis, we consider different parametric specifications of the function f to estimate the coefficient of variety.

We present our preliminary results in Table 2. We focus on the sample of impressions from the fourth to the tenth exposures in a session. The motivation behind our choice of the fourth exposure as the starting point is simply to have more variation in variety. We use the tenth exposure as the ending point only to ensure that we have enough observations per exposure number. In the first column, we consider the most basic model, in which we simply regress the outcome on variety, controlling for the ad and exposure number fixed effects. As we show in the first column of Table 2, the coefficient of variety is positive and statistically significant. In light of Figure 5, Panel B, we know that the assignment to variety is confounded by advertisers' targeting. Therefore, in the model in the second column, we control for all the targeting variables presented in Table 1. While the magnitude of the variety coefficient changes, the sign and significance remain unchanged.

Next, we focus on another potential confound: session length. Given the sequential nature of the variety assignment, some users may drop out in the middle of the session (i.e., the session length is not the same across all users). In particular, if their decision to leave is influenced by the variety assignment, this would create a dynamic selection issue. A simple (yet insufficient) solution would be to estimate the effects of variety with session length held constant. This ensures that we only compare users who made the decision to leave the session at the same point. In the third column of Table 2, we control for session length in addition to prior controls. The estimate shows the same pattern: the coefficient of variety remains positive and significant.

Finally, note that changes in variety within a session do not happen in isolation. Other characteristics of the session (that influence the click outcome) can covary with variety of previous ads. As such, without proper controls, the coefficient of variety may actually pick up the effects of these session-level variables rather than variety. Therefore, in the fourth column of Table 2, we add two session-level controls: (1) $Freq_{i,t}$, which is the number of prior exposures of the current ad ($A_{i,t}$) within the session, and (2) $Space_{i,t}$, which is temporal space between the current ad and the last time it was shown in the session. For example, if session i shows the sequence , we have $Freq_{i,5} = 2$ and $Space_{i,5} = 5 - 2 = 3$.[11] While the qualitative results on the coefficient of variety do not change, the significance level of coefficients for both $Freq_{i,t}$ and $Space_{i,t}$ highlights the importance of controlling for other session-level variables.[12]

Overall, our preliminary results in Table 2 show a strong statistical link between the variety of previous ads and the click outcome on the current ad. The patterns are robust: we find the same patterns when we use an entropy-based measure of variety (e.g., Shannon entropy) and when we use a logistic regression to estimate our binary outcome (see Web Appendix B).

Challenges in Establishing Causality

We now discuss why our preliminary analysis falls short of establishing a causal link between the variety of previous ads and the click outcome on the current ad. These reasons are related to three different aspects of our main variable, $V_{i,t}$.

[11] If the ad has not been shown in the session before, we have $Space_{i,t} = t$.

[12] These variables are correlated with variety as follows: $\rho(V_{i,t}, Freq_{i,t}) = -.1702$, and $\rho(V_{i,t}, Space_{i,t}) = .4168$.

Table 2. Preliminary Results on the Effect of Breadth of Variety on Click Outcome.

	Dependent Variable: Click $V_{i,t}$			
	(1)	**(2)**	**(3)**	**(4)**
Variety ($V_{i,t}$)	.00245***	.00107***	.00107***	.00084***
	(27.62)	(11.27)	(11.24)	(8.18)
$Freq_{i,t}$				−.00041**
				(−3.11)
$Space_{i,t}$				.00031***
				(4.78)
Ad FE	✓	✓	✓	✓
Exposure number FE	✓	✓	✓	✓
Targeting variables FE		✓	✓	✓
Session length FE			✓	✓
R^2	.007	.009	.009	.009
Adjusted R^2	.007	.008	.009	.009
Number of observations	1,993,542	1,993,542	1,993,542	1,993,542

**$p < .01$.
***$p < .001$.
Notes: FE = fixed effect.

First, as Figure 5, Panel B, illustrates, assignment to variety is not fully exogenous. While the model in the second column of Table 2 tries to address this issue by controlling for targeting variables through fixed effects for each targeting subcategory, there may still be a more complicated selection (e.g., through the interaction of different targeting variables). Thus, it is essential for our empirical framework to guarantee that, conditional on controls, assignment to variety is fully exogenous.

The second issue stems from the fact that the receipt of variety is different from assignment to it. Users may leave the session at any point they want. Therefore, discrepancy between the receipt of and assignment to variety can create identification challenges because we only observe the receipt of variety, whereas randomization (if any) happens at the assignment level. Note that our control for session length in the third column of Table 2 only helps when users decide how long they want to stay in a session before it starts. However, there may be more complex scenarios in which users' decision to leave is a function of their variety assignment. Thus, our empirical framework needs to account for the discrepancy between the receipt and assignment to variety.

Finally, variety is not well-defined as a treatment. It is difficult to isolate an exogenous increment in variety such that only variety changes one unit, with all else remaining constant. Thus, even with complete randomization of variety and no dynamic selection, the models in Table 2 estimate a composite effect of variety and other session-level variables that covary with variety. Controlling for other session-level variables (Column 4 in Table 2) helps with this issue, but we cannot verify whether these controls are sufficient. Thus, a primary goal in our empirical framework is to isolate the effects of variety to the extent possible.

In summary, these three issues preclude us from making causal statements based on our preliminary analysis. Next, we discuss our empirical strategy to address these issues.

Empirical Framework

Problem Definition

Before we formally define our problem, let us we restate the goal of our study: we want to examine the extent to which a random increment in the variety of previous ads changes the click outcome, holding everything else constant. As our preliminary analysis highlights, a fundamental challenge in achieving our goal is the difficulty in randomizing this increment in isolation: an increment in variety may change other session-level covariates, thereby violating the *ceteris paribus* interpretation. For example, an increment in the variety of previous ads can change the spacing between ads as well. The central challenge is whether we can achieve some level of separability between an increment in variety and the rest of the information in the prior sequence of ads shown so that we can manipulate variety, holding all else constant.

An interesting feature of our measure of variety that helps with this separability condition is the fact that an increment in this measure has a clear interpretation at the exposure level: every exposure that shows an ad that has not been shown before adds one unit to the breadth of variety, thereby capturing the event of "increase in ad variety." We can formalize this intuition by defining the binary variable $W_{i,t}$ for any $t \geq 3$ as follows:

$$W_{i,t} = 1(A_{i,t-1} \notin \{A_{i,1}, \ldots, A_{i,t-2}\}). \quad (4)$$

Here, $W_{i,t}$ takes the value of one if the ad shown in exposure $t-1$ is distinct from the set of ads shown in the prior $t-2$ exposures, thus increasing ad variety by one unit.[13] This allows us to derive a binary decomposition of variety at the

[13] For example, in a session i with ads, we have $W_{i,3} = 1$ and $W_{i,4} = 0$.

exposure level for any $t \geq 3$, as follows:

$$\begin{aligned} V_{i,t} &= |\{A_{i,1}, \ldots, A_{i,t-1}\}| \\ &= |\{A_{i,1}, \ldots, A_{i,t-2}\}| + 1(A_{i,t-1} \notin \{A_{i,1}, \ldots, A_{i,t-2}\}) \\ &= V_{i,t-1} + W_{i,t} \\ &= 1 + \sum_{s=3}^{t} W_{i,s}. \end{aligned} \tag{5}$$

The intuition behind this decomposition is simple: for each distinct ad in the set of prior ads, $W_{i,t}$ takes a value of one only once (the first time each distinct ad is shown in the sequence).

The recursive relationship $V_{i,t} = V_{i,t-1} + W_{i,t}$ in Equation 5 illustrates how we approach separability in our problem: focusing on the increment in the last exposure helps us isolate its effects from the sequence of ads shown before that. As such, at any exposure $t \geq 3$, we define the binary variable $W_{i,t}$ as the treatment variable of interest. An intuitive definition of $W_{i,t}$ is "an increase in ad variety" in the previous exposure. Our goal would then be to measure the effect of this treatment on the click outcome in the current exposure. Figure 6 visualizes our research design, in which the variety treatment is assigned and received in exposure $t - 1$ and the outcome is collected in exposure t.[14] This research design helps us overcome/avoid the two shortcomings in our preliminary analysis: (1) receipt of variety is the same as being assigned to it, and (2) the treatment increment has a clearer definition (i.e., it does not covary with session-level information up until the treatment stage [pretreatment period in Figure 6]).

With the goal of measuring the causal effects of "an increase in ad variety at point $t - 1$" on the "click outcome at point t," we write the main equation we want to estimate as follows:

$$\begin{aligned} Y_{i,t} &= \beta W_{i,t} + g(X_{i,t}, H_{i,t}) + \epsilon_{i,t}, \\ &= \beta W_{i,t} + \underbrace{g_{pre}(X_{i,t-2}, H_{i,t-2})}_{\text{pretreatment controls}} + \underbrace{g_{post}(A_{i,t-1}, A_{i,t}; H_{i,t-2})}_{\text{posttreatment controls}} + \epsilon_{i,t}, \end{aligned} \tag{6}$$

where β captures the effects of variety and the second equality represents our use of separability in this problem. We explicitly assume that we can additively separate our controls into two categories: (1) pretreatment controls that account for selection in the assignment to our treatment (i.e., $g_{pre}(X_{i,t-2}, H_{i,t-2})$), (2) posttreatment controls that help separate the treatment effects from other posttreatment variables that covary with our treatment (i.e., $g_{post}(A_{i,t-1}, A_{i,t}; H_{i,t-2})$).

Before we specify the functions g_{pre}, g_{post}, we describe the challenges that we face and our identification strategy to address these challenges. Consistent with our identification strategy, we then present our full model specification.

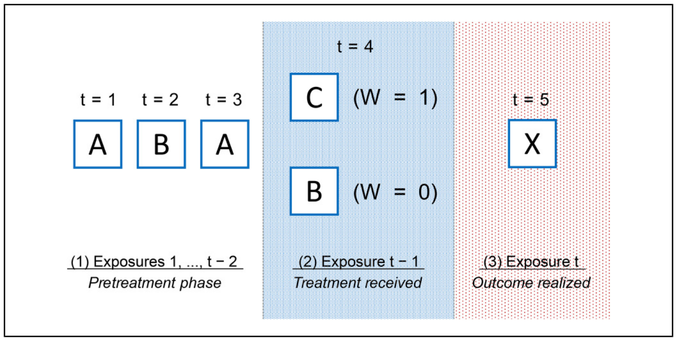

Figure 6. A visual depiction of our research design.
Notes: There are three phases: (1) the pretreatment period, which includes all the information about the user as well as all the ads seen up until exposure $t - 1$; (2) the treatment period in exposure $t - 1$, when the user is assigned to the variety condition or control; and (3) the outcome collection period t, in which we collect the click outcome.

Identification Strategy

We face three key challenges in estimating the treatment effects specified in Equation 6:

- *Pretreatment confounding:* Assignment to treatment is confounded by pretreatment variables. That is, a user's propensity to receive the treatment is a function of these pretreatment variables. Function g_{pre} in Equation 6 should address this type of confounding in our problem.
- *Dynamic selection (posttreatment censoring):* While being assigned to variety is equivalent to receiving it, the user may leave right after receiving the treatment (i.e., just after the $(t - 1)$th exposure), thereby censoring some of the posttreatment variables and the outcome. We need to address this dynamic selection problem in our identification strategy.
- *Posttreatment confounding:* The posttreatment phase is defined from treatment assignment to the outcome collection phase. During this time, other important variables also covary with our treatment (e.g., the identity of the ads shown, their recency, and their frequency). This imposes a challenge if we want to isolate the treatment effect of an increase in ad variety. Therefore, we need to control for any posttreatment confounding to isolate the effects of variety. Function g_{post} in Equation 6 controls for this type of confounding.

Solution to Pretreatment Confounding. The variation in our treatment variable $W_{i,t}$ can be confounded with the pretreatment variables. As Figure 3, Panel C, shows, advertisers favor more popular sessions whose characteristics are associated with higher CTR. As a result, assignment to treatment is more likely in more popular sessions simply because there are more

[14] A natural question is why we do not compare the outcomes at point $t - 1$. The main reason is that it would not be possible to separate the variety effects from the ad effects without assuming a certain specification. Under these specifications, the results are directionally the same as our main results. In the current research design, however, we have greater power to control for the ad effects in the period after treatment assignment.

ads in the inventory to increase the variety of prior sequence. For example, in our data, over 83% of the impressions in the largest province are assigned to the treatment condition; this percentage drops to 61% for impressions in a small province. Furthermore, as Figure 6 shows, assignment to treatment at exposure $t-1$ depends on the prior ad assignments in the session at exposures 1 to $t-2$. For example, if the variety of prior ads in a session is higher, the likelihood of being assigned to the treatment is lower by construction. For example, 76% of impressions with a prior variety of one received the treatment, whereas only 40% of impressions with a prior variety of four received the treatment. We can state this challenge as follows:

Challenge 1: There is nonrandomness in the treatment assignment. For an arbitrary exposure t in two random sessions i and j, the propensities of receiving the treatment are not necessarily the same; that is, $\Pr(W_{i,t}=1) \neq \Pr(W_{j,t}=1)$.

To solve this problem, we focus on the source of nonrandomness in the treatment variable. Given the definition of our treatment variable in Equation 4, we know that the distribution of the ad allocation process fully determines the distribution of treatment assignment. Thus, we focus on the ad allocation process as the source of nonrandomness in the treatment. In light of Equation 1, we know that advertisers' bids, quality scores, and participation decisions fully determine the ad allocation process. Therefore, these are the only three possible sources of nonrandomness. We use this observation to formally express the following proposition:

Proposition 1: The distribution of propensity scores for ad assignment $\pi_{i,t}(a)$ for any exposure/impression is only a function of impression-specific observables, $X_{i,t}$, in the data.

Proof. See Web Appendix C.1.□

This is a crucial result because it ensures that, conditional on exposure, there are no user- or impression-specific unobservables that affect ad assignment that are observable to advertisers (and the ad network) but not to the researcher. We now link Proposition 1 to the propensity scores for our treatment variable $W_{i,t}$ in the following remark:

Remark 1: Let $e(W_{i,t})$ denote the propensity score to be assigned to the treatment condition; that is, $e(W_{i,t}) = \Pr(W_{i,t}=1)$. Then, $e(W_{i,t})$ is only a function of impression-specific observables because $e(W_{i,t})$ is a linear function of $\pi_{i,t}(a)$ at any point; that is, $e(W_{i,t}) = \sum_{a \notin H_{i,t-2}} \pi_{i,t-1}(a)$.

It is important that $e(W_{i,t})$ is only a function of impression-specific observables because it shows the unconfoundedness of our treatment variable $W_{i,t}$. That is, for any set of potential outcomes $\mathcal{Y}_{i,t}$, we have $\Pr(W_{i,t}|X_{i,t-1}, H_{i,t-2}) = \Pr(W_{i,t}|\mathcal{Y}_{i,t}, X_{i,t-1}, H_{i,t-2})$.

Our approach to directly address Challenge 1 is to use one of the key results of Rosenbaum and Rubin (1983): it is sufficient to control for the propensity scores of the treatment variable ($e(W_{i,t})$) under unconfoundedness. As such, the challenge lies in estimating the propensity scores of the treatment variable using the pretreatment information. Estimation of $e(W_{i,t})$ is a prediction task in which we need to regress observed $W_{i,t}$ in the data on $X_{i,t-1}$ and $H_{i,t-2}$. Because this is a prediction task, we can use a machine-learning method that captures more complex relationships and achieves better predictive accuracy. We summarize our specific solution to address the pretreatment confounding challenge in three steps:

- *Step 1*: We estimate $e(W_{i,t})$ using an XGBoost model (Chen and Guestrin 2016). That is, $\hat{e}(W_{i,t}) = \widehat{XGB}(X_{i,t-1}, H_{i,t-2})$. For details of our propensity estimation approach, see Web Appendix D.1.
- *Step 2*: We assess covariate balance to confirm that the inverse propensity weight-adjusted (IPW-adjusted) distribution of each variable is reasonably similar across treatment and control groups. For details of our balance assessment, see Web Appendix D.2.
- *Step 3*: We feed the inverse propensity weights into the regression model to account for the pretreatment confounding issue.

Solution to Dynamic Selection. As we discussed previously, users' decision to leave a session can cause a missing data problem. In our setting, dynamic selection only happens at the outcome level: while all users available at point $t-1$ receive the variety treatment, their outcome at exposure t may be missing if they decide to leave the session right after the treatment. This type of dynamic selection does not create estimation bias if users' decision to leave is completely random. However, dynamic selection can interfere with inference if a user's decision to leave the session is (1) a function of the variety assignment in exposure $t-1$ or (2) the user's own characteristics (especially unobserved) that also affect their click probability.[15]

Figure 7 uses a simple example to illustrate the general form of dynamic selection in our problem and how it can interfere with inference. In this example, there are six users who have seen ads in their first three exposures. These users are then randomly assigned to treatment and control conditions in the treatment state ($t=4$). In all these cases, the user is supposed to see ad X in the fifth exposure. However, only three of the six users were exposed to the fifth exposure: one in the treatment condition and two in the control condition. To understand how dynamic selection can interfere with inference, suppose that only one impression in each condition

[15] Dynamic selection has also been recognized in other settings in which users' survival is a function of user-level unobservables. For example, Yoganarasimhan (2013) accounts for persistent user-level unobservables within dynamic structural models by explicitly incorporating unobserved heterogeneity in users' utility functions and state transitions.

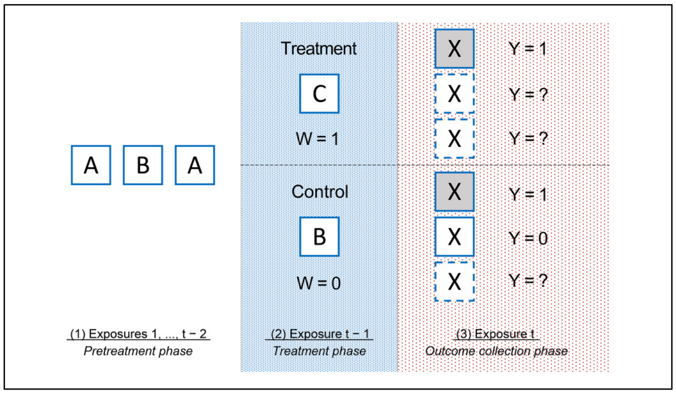

Figure 7. Example of posttreatment censoring.
Notes: The dotted impressions are exposures that did not occur because the user left the session before their occurrence, and the gray shaded impressions denote clicks.

has been clicked. If we observe all the exposures (i.e., dynamic selection is not an issue), then we would correctly infer that there is no difference in CTRs across the treatment and control conditions. However, if we rely solely on the observed data, we would infer a CTR of 1 for the treatment condition and a CTR of (1 / 2) for the control condition. This would lead to the incorrect inference that treatment results in a higher CTR.

To address this issue, we need to impute the missing impressions. In particular, we need to impute two characteristics of these missing impressions: (1) the ad the user would have been assigned to and (2) the corresponding click outcome. Let $A^*_{i,t}$ and $Y^*_{i,t}$ denote the ad and click outcome for both observed and missing impressions. For observed impressions, we have $A^*_{i,t}=A_{i,t}$ and $Y^*_{i,t}=Y_{i,t}$. We can now formally state this challenge as follows:

Challenge 2: For a user who has received the treatment at exposure $t-1$ but did not stay for exposure t, we need to impute the set of $\{A^*_{i,t}, Y^*_{i,t}\}$, where $A^*_{i,t}$ is the ad this user would have been assigned to in exposure t and $Y^*_{i,t}$ is the corresponding outcome.

We first discuss our imputation strategy for missing ads. Let τ denote the exact time stamp of an exposure in session i.[16] Next, let the ad shown in session i at time stamp τ be drawn from the distribution $\mathcal{A}_i(\tau)$. Then, it is easy to show the following:

Proposition 2: For any two exposures in sessions i and j with the same targeting characteristics, the distribution of ad allocation is the same at any arbitrary time stamp τ; that is, $\mathcal{A}_i(\tau) \equiv \mathcal{A}_j(\tau)$.

Proof. See Web Appendix C.2.□

Proposition 2 is a direct result of the ad allocation process in Equation 1. Given this proposition, we can use the actual ad assignment in exposures from other sessions that are not part of our sample (but share the same targeting characteristics) to impute the intended ad assignment for exposures in the sessions in our sample.[17] We present the details of our imputation approach in Web Appendix E. Note that unlike most imputation approaches that use models to approximate the original distribution and simulate missing data from this approximate distribution, our approach is model-free and guarantees that the imputed exposures are drawn from the original distribution.[18]

Finally, we impute the missing outcomes as zero simply because the user is not available to click on the ad. An alternative approach would be to impute the outcome as the click decision the user would have made if he or she had stayed in the session (Little and Yau 1996). While this is the conventional approach in medical studies, we believe that our approach is the right one for our context because the user's decision to leave prevents the event in which the outcome of interest happens (the user clicks on the next ad).[19] Nevertheless, we run a series of robustness checks to show that our results are not driven by this modeling choice (see Web Appendix F.4).

Solution to Posttreatment Confounding. While unconfoundedness rules out pretreatment confounding, the nature of our treatment gives rise to the issue of posttreatment confounding. That is, from the point users are assigned to the treatment (at exposure $t-1$) to the point we collect the outcome, it is not just the treatment assignment that is different across treatment and control groups; other information about exposures $t-1$ and t may also differ, such as the specific ads shown as well as their prior frequency and spacing in these exposures. This brings us to our third challenge:

Challenge 3: There is a function $g_{post}(A_{i,t-1}, A_{i,t}; H_{i,t-2})$ that is defined based on posttreatment inputs and is correlated with $W_{i,t}$ and $Y_{i,t}$. Therefore, failure to control for this function leads to omitted-variable bias.

To address this challenge, we need to specify function g_{post} such that it captures any posttreatment variable in that is correlated with both $W_{i,t}$ and $Y_{i,t}$. To simplify the problem, we need to define variables that capture the relationship between $A_{i,t}$ and $A_{i,t-1}$ with the past sequence $H_{i,t-2}$. We focus on two sequence-dependent variables that have been shown to drive ad effects: frequency and spacing of the ad shown in an exposure in the

[16] Note that time stamp τ is distinct from exposure number t; τ is the exact time at which an impression occurs; for example, if the first impression in a session occurred at 9:21:34 P.M. of a specific day, then τ is 9:21:34 P.M., whereast = 1.

[17] We do not use the sessions in our sample for imputation, because doing so can cause finite-sample issues in some parts of the data.

[18] While we use this specific approach to overcome dynamic selection, the general solution is to approximate the allocation distribution from the data, which is feasible under the unconfoundedness of ad allocation.

[19] This is different from medical studies in which the outcome is often an objective measure of one's health rather than a choice.

session. For any exposure t, we define the within-session frequency of the ad shown at this exposure, $Freq_{i,t}$, as follows:

$$Freq_{i,t} = \sum_{s=1}^{t-1} 1(A_{i,s} = A_{i,t}). \tag{7}$$

This variable measures the number of times the ad shown at exposure t has been shown earlier in the session. We define $Space_{i,t}$ as the spacing between the ad shown at exposure t and the last time this ad was shown (if any):

$$Space_{i,t} = t - \max\{s \cup \{0\} | A_{i,s} = A_{i,t}\}, \tag{8}$$

where spacing is defined in terms of exposure numbers and $Space_{i,t} = t$ if the ad shown at t was not shown before. We use these two variables to eliminate dependence on the past. That is, we assume that all the information in $A_{i,t-1}$, $A_{i,t}$, $H_{i,t}$ is summarized in $A_{i,t-1}$, $Freq_{i,t-1}$, $Space_{i,t-1}$, $A_{i,t}$, $Freq_{i,t}$, and $Space_{i,t}$. Given this assumption, we need to include all six variables that are not perfectly colinear with our treatment $W_{i,t}$. Note that both $Freq_{i,t-1}$ and $Space_{i,t-1}$ are perfectly colinear with our treatment:

$$\begin{aligned} W_{i,t} &= 1(A_{i,t-1} \notin |\{A_{i,1}, \ldots, A_{i,t-2}\}|) = 1(Freq_{i,t-1} = 0) \\ &= 1(Space_{i,t-1} = t - 1). \end{aligned} \tag{9}$$

Therefore, we exclude these two variables from our model. However, we return to these two variables when we explore the potential behavioral mechanisms.

In summary, we include the following four variables to control for posttreatment confounding: $A_{i,t-1}$, $A_{i,t}$, $Freq_{i,t}$, and $Space_{i,t}$. Together, our posttreatment controls allow us to isolate the effects of treatment to the maximum extent possible. In Web Appendix F.2, we conduct a series of robustness checks to confirm that our findings are robust to alternative estimation approaches.

Model Specification

To address the pretreatment confounding issue, we use IPW-adjusted linear regression in which we weight impressions by their inverse propensity score of receiving treatment because it is an efficient estimator when using propensity scores to estimate treatment effects (Hirano, Imbens, and Ridder 2003). Next, to account for dynamic selection, we run our regression using the fully imputed variables for the ad and click outcome (i.e., $Y^*_{i,t}$ and $A^*_{i,t}$). Finally, we address posttreatment confounding by controlling for $A_{i,t-1}$, $A_{i,t}$, $Freq_{i,t}$, and $Space_{i,t}$. Thus, the main version of our model is the following IPW-adjusted regression specification:

$$\begin{aligned} Y^*_{i,t} =& \beta W_{i,t} + \sum_q \gamma_q 1(Freq_{i,t} = q) \\ &+ \sum_s \delta_s 1(Space_{i,t} = s) + \alpha_0(A^*_{i,t}) + \alpha_1(A_{i,t-1}) + \zeta_t + \epsilon_{i,t}, \end{aligned} \tag{10}$$

where β captures the treatment effect; γ_q and δ_s are the coefficients for levels q and s of $Freq_{i,t}$ and $Space_{i,t}$, respectively; $\alpha_0(A^*_{i,t})$ and $\alpha_1(A_{i,t-1})$ control for the fixed effects of ads shown in exposures t and $t-1$; and ζ_t controls for exposure number fixed effects. We use this model as the main specification, but we also consider other specifications with more controls in the next section.

Results

Main Effects of Variety

Results from the main specification. We start by estimating the average treatment effect for our main specification in Equation 10 and present the results in the first column of Table 3. We use the sample of impressions from exposures $t = 4$ to $t = 10$.[20] Because we use IPW-adjusted regression, we use robust standard errors for inference. The positive and significant treatment coefficient indicates a positive causal link between an increase in the variety of prior ads and the click outcome. That is, showing an ad that increases variety in the sequence of ads results in a higher CTR on the next ad, holding all else constant. Our main finding highlights a source of externality in this market: an intervention in a given period affects outcomes in future periods. This contrasts with the common assumption in the online advertising marketplace, where each impression is sold as an independent unit.[21]

To interpret the magnitude of our coefficients, we compare them with the baseline CTR in the system. The baseline CTR for our sample from $t = 4$ to $t = 10$ is .0134, which means that approximately 1.34% of the impressions in our sample get clicked. The treatment coefficient in the first column of Table 3 is .00186. As such, the magnitude of our treatment coefficient accounts for 13.88% of the baseline CTR in our sample. This implies that an increase in ad variety can shift CTR by approximately 13.88%, holding all other variables fixed.

Next, we try other specifications by dropping some of the necessary controls. First, in the second column of Table 3, we run an unweighted least squares regression to compare the estimates with and without IPW adjustment. As we expected, both the magnitude and the statistical significance increase because the unweighted model does not account for the differences in propensity scores, thereby capturing the endogenous variation in treatment assignment. In the third column, we run a model without accounting for the dynamic selection issue. That is, we focus only on the sample of impressions that survived and drop the impressions in which the user left the session right after the treatment assignments. Note that the outcome for all the excluded impressions is zero. As a result, the estimated treatment coefficient is not directly comparable with the other coefficients in Table 3 because the

[20] This is because we want to start from an exposure that has a relatively high propensity of assignment for both treatment and control groups. For example, starting from $t = 2$ gives us very low propensity scores for the control condition. For the ending point, we want to end at a t that still has enough impressions. Note that our results do not change if we include all time periods.

[21] The common practice of running second- or first-price auctions to sell digital ads is based on the assumption that impressions are independent units.

Table 3. Average Effects of the Variety Treatment on the CTR.

	Dependent Variable: Click $V^{*}_{i,t}$			
Treatment ($\mathbf{W}_{i,t}$)	**(1) .00186*** (9.01)**	**(2) .00203*** (11.14)**	**(3) .00247*** (9.91)**	**(4) .00235*** (11.81)**
IPW adjustment	✓		✓	✓
Imputed sample	✓	✓		✓
Exposure (t) FE	✓	✓	✓	✓
$Freq_{i,t}$ indicators	✓	✓	✓	
$Space_{i,t}$ indicators	✓	✓	✓	
$A^{*}_{i,t}$ FE	✓	✓	✓	✓
$A_{i,t-1}$ FE	✓	✓	✓	✓
Number of observations	2,405,695	2,405,695	1,993,542	2,405,695
R^2	.006	.006	.007	.005
Adjusted R^2	.005	.006	.007	.005

***$p < .001$.
Notes: Numbers reported in parentheses are t-statistics computed on the basis of robust standard errors. FE = fixed effect.

samples are systematically different. To adjust for this, we need to multiply the coefficient in the third column by the ratio of two samples 1,993,542/2,405,695, which is .8287. The adjusted coefficient is .00205, which implies that we would overestimate the effects of treatment if we do not account for dynamic selection caused by the effect of treatment on users' decision to leave. This is because an increase in ad variety likely comes with a greater user propensity to leave the session, as an increase in ad variety can be perceived as increased ad load, which has been shown to have negative effects on usage in the prior advertising literature (Wilbur 2008). Finally, in the fourth column of Table 3, we drop the $Freq_{i,t}$ and $Space_{i,t}$ controls. Given that both these covariates are correlated with our treatment and likely associated with the click outcome, we expect a change in the treatment effect estimates. We find that we would overestimate the treatment effects if we do not control for these two covariates.

In summary, the results in Table 3 establish the main positive effects of an increase in ad variety on the click outcome on the next ad. The results also show the importance of controlling for all three types of confounding discussed in the article.

Robustness checks on the main effects. We perform a series of robustness checks on our main results. We discuss these models briefly here and refer readers to Web Appendix F for details.

First, in Table 3, we user a linear probability model for a binary outcome. In Web Appendix F.1, we present the results of a logistic regression for the same model specifications. Second, we consider models with overly conservative controls for both pre- and posttreatment variables. These models separately control for (1) the interaction of all targeting variables,[22] (2) user and hour-day fixed effects, (3) session fixed effects, and (4) different interactions of all posttreatment variables. As Web Appendix F.2 shows, our results consistently show the main effect: an increase in ad variety at any exposure results in higher CTR on the next ad.

Most notably, we employ an exact-matching approach to fully isolate our treatment effects from any pre- or posttreatment covariates. We match impressions based on the exact sequence of ads shown in the session, except for the ad shown in the treatment phase (exposure $t-1$). That is, two impressions belong to the same matching group if $(A_{i,1}, \ldots, A_{i,t-2}, A_{i,t})$ is exactly the same for them, and they only differ in $A_{i,t-1}$. We also control for the fixed effects of $A_{i,t-1}$ and the propensity scores of the treatment. Although our statistical power is substantially compromised in this case, our main findings still hold, and the results of this exact-matching approach show a significant and positive treatment coefficient (for more details on our exact-matching practice, see Web Appendix F.3).

Finally, we run a series of additional checks to establish the robustness of our results to (1) alternative approaches to imputation (see Web Appendix F.4), (2) different levels of clustering in standard errors (see Web Appendix F.5), and (3) a placebo treatment definition to ensure the data structure does not drive our results (see Web Appendix F.6). All these robustness checks confirm the validity of our main results.

Mechanism for the Effects of Variety

Theoretical underpinnings of the mechanism. In the previous section, we note that showing a new (or previously unseen) ad at $t-1$ increases the user's probability of clicking on the ad shown at t. To pin down the mechanism, we focus on the main feature that differs across our treatment and control groups, namely, the novelty of the ad shown at the treatment phase. A unifying result that emerges from both early and recent work in the behavioral literature is that novelty of stimuli shown in a given space increases people's attention to

[22] This approach is similar to Nair et al. (2017), who use firms' targeting decisions to control for the selection caused by targeting. While our main approach is based on using propensity scores, here we add the interaction of all targeting variables to make sure that our main results are not driven by our propensity score estimates.

that space (Han and Marois 2014; Helson 1948; Kahneman 1973). In our context, this means that increasing ad variety by showing a novel ad leads to higher attention to the advertising slot, thereby increasing consideration of and click probability on the next ad. We propose this attention-based account as the underlying mechanism for the variety effects we report.

There are three key advantages to using the aforementioned attention-based account. First, it is parsimonious because it only uses a well-established finding that users pay more attention to novel stimuli (in this case, ads) that have been used less recently and less frequently in the past. This account traces its roots to the early work on adaptation-level theory (Helson 1948), which has served as a theoretical foundation in the study of the variety effects (Redden 2008). Second, the foundation of our theory (i.e., users pay more attention to novel stimuli) is consistent with prior work in advertising and eye-tracking (Pieters, Rosbergen, and Wedel 1999), thereby providing contextual validity. Finally, our behavioral account delivers concrete predictions that are empirically testable. We can focus on different slices of the data where we expect to have a higher (or lower) gap in novelty between the treatment and control and test if the estimates consistently change with this gap in novelty. In the rest of this section, we adopt this strategy; that is, we make theory-driven predictions based on our proposed mechanism and then examine if they are empirically true.

Treatment effects across different control groups. Recall Equation 9:

$$\begin{aligned} W_{i,t} &= 1(A_{i,t-1} \notin |\{A_{i,1}, \ldots, A_{i,t-2}\}|) = 1(\text{Freq}_{i,t-1} = 0) \\ &= 1(\text{Space}_{i,t-1} = t-1). \end{aligned}$$

We know that an impression belongs to the treatment condition if the ad shown in period $t-1$ has not been shown in the past; that is, $\text{Freq}_{i,t-1} = 0$, and $\text{Space}_{i,t-1} = t-1$. Conversely, an impression belongs to the control condition if $\text{Freq}_{i,t-1} \neq 0$ and/or $\text{Space}_{i,t-1} \neq t-1$. In other words, our control condition constitutes a range of values for $\text{Freq}_{i,t-1}$ and/or $\text{Space}_{i,t-1}$. However, our underlying mechanism suggests that the exact levels of past frequency and spacing of the control condition matter. Specifically, we expect the control group to be less novel if the ad shown in the treatment phase has been shown more frequently (higher $\text{Freq}_{i,t-1}$) and/or more recently (lower $\text{Space}_{i,t-1}$). Building on these ideas, we offer the following concrete predictions:

Prediction 1: The treatment effect is higher when we compare the treatment group ($W_{i,t} = 1$) with a control group when the ad shown in the treatment phase has higher frequency. That is, as we increase $k > 0$ in $\text{Freq}_{i,t-1} = k$ to define the control group, the treatment effect increases.

Prediction 2. The treatment effect is higher when we compare the treatment group ($W_{i,t} = 1$) with the control group when the ad shown in the treatment phase has lower spacing (higher recency). That is, as we increase $l < t-1$ in $\text{Space}_{i,t-1} = l$ to define the control group, the treatment effect decreases.

To empirically test these predictions, we first partition the control group in our data into subgroups based on the frequency of the ad shown at $t-1$ (e.g., $W_{i,t} = 0$ and $A_{i,t-1}$ has been shown k times before). Then, we separately estimate the treatment effect against each of these control groups and present the results in Figure 8, Panel A. We observe an increasing trend in treatment effects as the frequency of the ad at $t-1$ increases. This suggests that when the last ad shown before t (i.e., $A_{i,t-1}$) was repeated several times earlier in the session (i.e., it is less novel), users pay less attention to the current ad. We perform a similar exercise by partitioning the control group into different subgroups based on recency or spacing (e.g., $W_{i,t} = 0$ and $A_{i,t-1}$ has been shown l exposures before) and present the results in Figure 8, Panel B. Here, we observe a decreasing pattern; specifically, the treatment effects decrease as the spacing of the ad shown in the control condition increases. The highest treatment effect occurs when we compare the treatment condition with the control condition that has repeated the ad before ($A_{i,t-1} = A_{i,t-2}$), which is equivalent to a spacing level of one. Interestingly, when we increase the spacing level to six in the control condition (i.e., $A_{i,t-1}$ was shown six impressions prior to $t-1$), we no longer observe a significant treatment effect. This suggests that repeating an ad that was shown much earlier in the session is almost as good as showing a new ad (i.e., the treatment condition in which variety increases by one unit). Together, these findings provide support for our proposed mechanism. For details of the regression used in Figure 8 and additional robustness checks on these findings, see Web Appendix G.1.

Heterogeneity across usage frequency and recency. To further explore the idea that user response to the variety treatment is driven by stimulus novelty and attention, we focus on two user-level features that capture a user's presession exposure to ads: (1) *usage frequency*, or the number of ad impressions the user has seen in prior sessions, and (2) *usage recency*, or the time lapse between the start of the current session and the end of the user's previous session. This variable captures how long ago the user was exposed to an ad (before the current session): when the gap is short, usage recency is high.

The theoretical mechanism we have proposed suggests that it is more difficult to shift users' attention if they have seen more ads in the past. Thus, we expect to observe higher treatment effects for users with a lower frequency of prior ads. Similarly, users who had more recent interactions with ads are less likely to be responsive to our treatment because their memory of some ads may be fresh/recent. As such, we expect the within-session interventions to be less effective in shifting users' attention when they are in the high-recency condition. Together, we offer the following two predictions based on our mechanism:

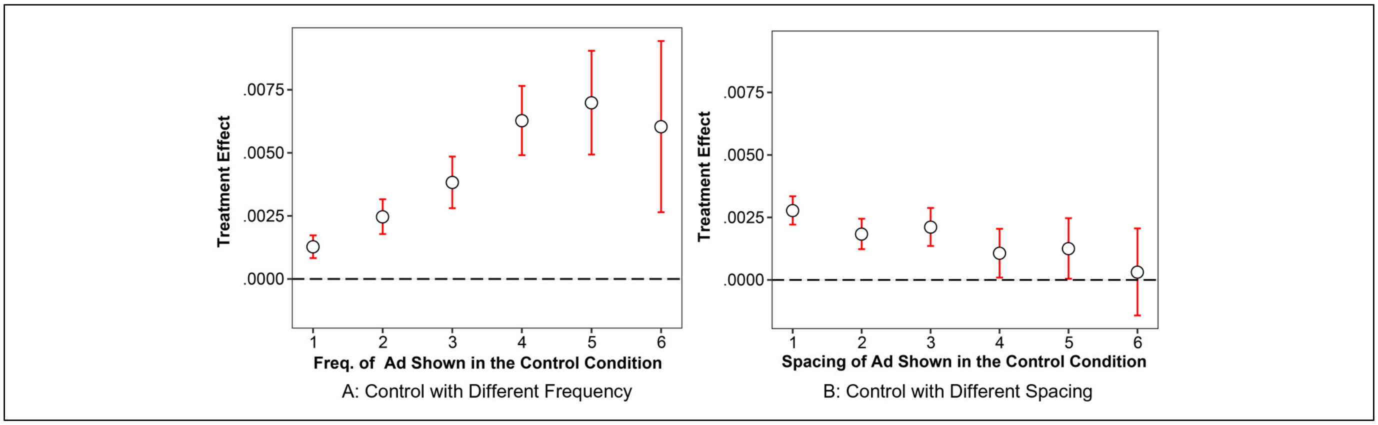

Figure 8. Treatment effects compared with different control groups defined by the frequency ($Freq_{t-1}$) and spacing ($Space_{t-1}$) of the ad shown in the control condition.
Notes: Confidence intervals are built using the robust standard errors from the IPW-adjusted regression model.

Prediction 3. The treatment effect is higher for users with low usage frequency than for users with high usage frequency.

Prediction 4. The treatment effect is higher for users with low usage recency than for users with high usage recency.

We empirically test these predictions in our data. We perform a rough median split and define the high- (low-) usage-frequency sample as the set of impressions where the user has seen more (less) than 100 impressions in prior sessions. We then estimate the treatment effects separately for these two partitions of the data. Next, we perform a similar exercise for usage recency, where one hour is a rough median split. We then use one hour as a threshold to partition impressions into low- and high-recency buckets and estimate separate treatment effects for each bucket. The estimated treatment effects from these analyses appear in Figure 9. Note that the treatment effects are significant and positive only in low-usage-frequency or low-recency conditions. When usage frequency or recency is high, the treatment is statistically insignificant.[23] These findings are consistent with Predictions 3 and 4, which are based on the attention-based mechanism we have proposed.

Heterogeneity across past variety. We now examine heterogeneity in our treatment effects across past variety. Given our mechanism, we expect the treatment to be less effective if prior variety is already high. This is because the increment in attention from treatment assignment would be lower when attention level is already high due to higher past variety. We state this prediction as follows:

[23] For robustness, we check if the results are significant in a regression with interaction terms and find the same patterns (for details, see Web Appendix G.2). We further investigate the source for this heterogeneity and document time-varying user characteristics (e.g., same user when they haveseen few vs. many ads) as the main source.

Prediction 5: The treatment effect is lower when the variety of is higher.

We test this prediction by estimating models that also include an interaction of the treatment with past variety. We consider three different measures of past variety in our analyses: (1) $V_{i,t-1}$, which is simply the number of distinct ads shown in the first $t-2$ exposures; (2) $\log(V_{i,t-1})$, which is the logarithm of the previous measure; and (3) $V_{i,t-1} / (t-1)$, which normalizes our first measure across exposure numbers. The results from this exercise appear in Table 4. Across all specifications and definitions of prior variety, our interaction term is negative and statistically significant. This implies that the treatment effect is reduced as past variety increases; that is, a novel ad (or the variety treatment) is less likely to increase users' attention when their attention level is already high as a result of seeing a higher variety of novel ads in prior exposures.

Implications

Managerial Relevance of Variety

Digital platforms have the ability to deliver numerous interventions to their user base. This inherently makes variety an important construct because users are exposed to many different marketing interventions every day. In addition, platforms increasingly engage in activities whose byproduct is more exploration and increased variety of interventions. These activities include (1) using adaptive exploration/exploitation algorithms to learn about consumer taste more efficiently (Lattimore and Szepesvári 2020); (2) adopting different fairness criteria to achieve greater parity across different demographic groups, which induces more randomization in ad allocation and thereby ad variety (Dwork et al. 2012); (3) employing algorithms that operate on the basis of increased diversification and randomization to prevent polarization (Celis et al. 2019); (4) building algorithms to enhance the reachability of recommender systems, which ensures that the

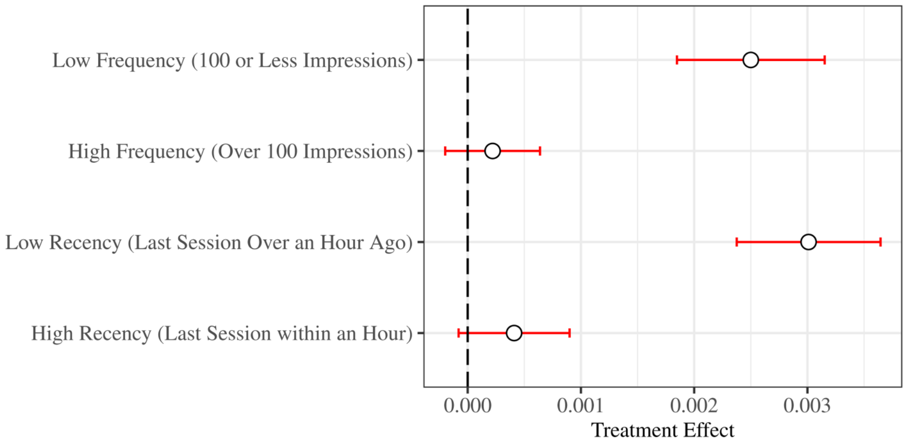

Figure 9. Heterogeneity in variety effects across usage frequency and recency.
Notes: Confidence intervals are built using robust standard errors in the IPW-adjusted regression.

Table 4. Heterogeneity in the Effects of Variety Across Past Variety.

	Dependent Variable: Click ($Y^*_{i,t}$)		
	(1) Past variety $= V_{i,t-1}$	(2) Past variety $= \log(V_{i,t-1})$	(3) Past variety $= \frac{V_{i,t-1}}{t-1}$
Treatment ($W_{i,t}$)	.00325***	.00374***	.00343***
	(7.18)	(5.89)	(7.69)
Past variety	.00156***	.00734***	.00459***
	(11.31)	(10.83)	(11.66)
Treatment ×past variety	−.00033**	−.00210**	−.00111**
	(−2.59)	(−2.67)	(−2.93)
Controls in Equation 10	✓	✓	✓
Number of observations	2,405,695	2,405,695	2,405,695
R^2	.006	.006	.006
Adjusted R^2	.005	.005	.005

** $p < .01$.
*** $p < .001$.

recommender system does not systematically make some of the items out of reach (Dean, Rich, and Recht 2020); and (5) committing to greater diversity standards in advertising. As these activities increase the variety of marketing interventions, it becomes increasingly important for managers and platforms to understand the downstream consumer-level consequences of this variety.

Takeaways for Platforms and Advertisers

From the platform's perspective, eyeballs or ad impressions are valuable resources, and increasing users' attention and clicks leads to higher revenue for them (especially in CPC settings). This explains why major platforms such as Google or Facebook invest heavily in research groups that help build better CTR prediction machines (He et al. 2014; McMahan et al. 2013). At a high level, the current research offers new insights into the CTR prediction problem for platforms by recognizing the causal effects of a new construct of ad variety that helps platforms improve their CTR prediction algorithms.

However, the role of ad variety goes beyond just improving CTR prediction machines; it causes an important externality that has implications for auction design and monetization of ad impressions. Our findings suggest that an increase in ad variety at one exposure changes the likelihood of clicks on future exposures, which violates the assumption of independence of impressions made in commonly used mechanisms such as first- or second-price auctions (for well-known examples in the digital advertising context, see Edelman, Ostrovsky, and Schwarz 2007; Varian 2007). In light of this externality, it is neither efficient nor optimal to sell an impression to the highest-bidding ad, as other competing ads may create greater positive externalities through increasing ad variety. Thus, it is essential for platforms to develop auctions that incorporate such externalities. In a recent paper, Rafieian (2020) presents a revenue optimal dynamic auction mechanism that accounts for the interdependence of impressions and quantifies the loss in platform revenues when interdependence is ignored.

Our findings also have implications for platforms that use adaptive experimentation tools such as contextual bandits to decide which treatment (e.g., ad copy or promotional content) to show at a given exposure. These approaches often assume the independence of rewards across treatment arms. However, our findings challenge this assumption and highlight the need to develop more dynamic experimentation approaches (Rafieian 2022; Theocharous, Thomas, and Ghavamzadeh 2015).

Finally, while we view this problem through the lens of a platform, our research also has implications for advertisers. First, our results suggest that advertisers may benefit from showing a variety of creatives, thereby helping advertisers trade-off repeating and varying creatives in their ad campaigns. Second, advertisers can incorporate information about the effects of ad variety into their decision making. Although past variety may often be unobserved by advertisers, larger demand-side platforms that bid on behalf of multiple ads can better incorporate our findings. However, these implications for advertisers must be interpreted with the necessary caveat that we study click as the main outcome of interest, not conversion.

Attention-Based Measures as a Potential Solution

While the foregoing discussion highlights the challenges caused by externalities due to variety effects, our analysis also provides directions for some solutions. In light of our attention-based mechanism, platforms can develop attention-based measures of the form $\lambda_{i,t}$ that are defined at the impression level and capture the past frequency and spacing of ad interventions; that is, $\lambda_{i,t} = \sum_{s=1}^{t-1} f_t(\text{Freq}_{i,s}, \text{Space}_{i,s})$.[24] Such attention-based measures can be used in different ways. First, platforms can provide this information as a targeting tool to advertisers, which can resolve the externality issue through market equilibrium. Second, platforms can use these attention-based measures to approximate the externality that an impression would impose on future impressions and use that to modify the standard first- and second-prices auctions. Third, these measures can be used as contexts in contextual bandits to mitigate the aforementioned issues with these algorithms in light of the effects of ad variety.

Applicability of the Methodological Framework

Within the advertising domain, platforms can use our framework as long as the unconfoundedness of ad allocation is satisfied. While this condition is satisfied in a standard field experiment, a full experiment is not necessary. It can also be easily satisfied by platforms if they can incorporate a small amount of randomization in their ad allocation mechanism (without significantly hurting their revenues). Many platforms already implement such approaches by adopting ϵ-greedy policies that select the optimal action by $1 - \epsilon$ probability but give a nonzero probability to all other actions (Theocharous, Thomas, and Ghavamzadeh 2015). An alternative approach is to randomize allocation only for a small portion of their total traffic; some platforms such as Bing use this approach (Ling et al. 2017). Overall, platforms that employ such randomization practices can easily adopt our framework.

Finally, our framework is applicable to domains other than advertising, where one may want to study the impact of an increase in variety/diversity. This includes studies at the intersection of digitization and diversity in sequential settings. For example, our framework could be used to study how increased diversity in music consumption affects consumer behavior in music-streaming channels or to examine how app users respond to an increased variety of push notifications. More broadly, our method can help in measuring the effect of treatments that are defined as a function of past behavioral history —that is, Treatment = f(Behavioral History). These interventions are increasingly relevant as platforms deliver more personalized interventions based on users' past behavior.

Conclusion

In many mobile and digital settings, users are often exposed to a sequence of short-lived marketing interventions within a short period of time. This is particularly true in the context of mobile in-app advertising, in which platforms use refreshable ad slots. Users are shown a sequence of potentially different ads within a session and therefore can be exposed to a large variety of ads within the same app-usage session. In this article, we examine how an increase in the variety of ads shown in a session affects user response to the next ad. We use the quasi-experimental variation in ad assignment in our data and propose a methodological framework that allows us to isolate the effects of an increase in ad variety. We apply our framework to data from the leading in-app ad network from an Asian country and empirically show that an increase in ad variety increases the CTR on the next ad by approximately 13%, holding everything else constant. We then explore the behavioral mechanism underlying this effect and examine an attention-based account based on prior behavioral literature: a novel ad that has been shown less frequently and less recently drives more attention to the advertising slot, thereby generating a higher CTR on the next ad. We test various predictions related to this behavioral account and find empirical evidence consistent with these predictions. Finally, we discuss the implications of our findings for managers and platforms.

Our article has certain limitations that present fruitful directions for further research. First, our results establish the effects of variety in a mobile in-app advertising setting. Future studies could extend our results to other advertising or nonadvertising contexts. Second, we study the problem from a platform perspective. Future work might adopt an advertiser-focused perspective and examine the role of ad variety from this point of

[24] There are also more direct approaches to measure attention with tracking technologies (McGranaghan et al. 2021).

view. Finally, we postulated the attention-based mechanism after observing the main effects of variety; as such, the results should be interpreted as confirmatory evidence of one possible mechanism. Our analysis does not rule out other possible mechanisms; nor do we conduct extensive theory testing by first proposing theories and then examining their applicability. Future work may benefit from exploring alternative theories for variety effects.

Acknowledgments

The authors are grateful to an anonymous firm for providing the data and to the UW-Foster High Performance Computing Lab for providing us with computing resources. They thank Ryan Dew, Simha Mummalaneni, and Hoori Rafieian for detailed comments that have improved the article. They also thank the participants of the 2018 UW/UBC Conference, 2018 ISMS Marketing Science Conference, and 2018 SICS Conference for their feedback.

Associate Editor

Kenneth Wilbur

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

Aravindakshan, Ashwin, Kay Peters, and Prasad A. Naik (2012), "Spatiotemporal Allocation of Advertising Budgets," *Journal of Marketing Research*, 49 (1), 1–14.

Arnosti, Nick, Marissa Beck, and Paul Milgrom (2016), "Adverse Selection and Auction Design for Internet Display Advertising," *American Economic Review*, 106 (10), 2852–66.

Braun, Michael and Wendy W. Moe (2013), "Online Display Advertising: Modeling the Effects of Multiple Creatives and Individual Impression Histories," *Marketing Science*, 32 (5), 753–67.

Bruce, Norris I., B.P.S. Murthi, and Ram C. Rao (2017), "A Dynamic Model for Digital Advertising: The Effects of Creative Format," *Message Content, and Targeting on Engagement. Journal of Marketing Research*, 54 (2), 202–18.

Bucklin, Randolph E. and Paul R. Hoban (2017), "Marketing Models for Internet Advertising," in *Handbook of Marketing Decision Models*, Berend Wierenga and Ralf van der Lans, eds. Berlin: Springer, 431–62.

Celis, L. Elisa, Sayash Kapoor, Farnood Salehi, and Nisheeth Vishnoi (2019), "Controlling Polarization in Personalization: An Algorithmic Framework," in *Proceedings of the Conference on Fairness, Accountability, and Transparency. New York: Association for Computing Machinery*, 160–69.

Chae, Inyoung, Hernán A. Bruno, and Fred M. Feinberg (2019), "Wearout or Weariness? Measuring Potential Negative Consequences of Online Ad Volume and Placement on Website Visits," *Journal of Marketing Research*, 56 (1), 57–75.

Chen, Tianqi and Carlos Guestrin (2016), "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: Association for Computing Machinery, 785–94.

Choi, Hana and Carl F. Mela (2019), "Monetizing Online Marketplaces," *Marketing Science*, 38 (6), 948–72.

Danaher, Peter J. (1995), "What Happens to Television Ratings During Commercial Breaks?" *Journal of Advertising Research*, 35 (1), 37–37.

Danaher, Peter J. and Guy W. Mullarkey (2003), "Factors Affecting Online Advertising Recall: A Study of Students," *Journal of Advertising Research*, 43 (3), 252–67.

Danaher, Peter J. and Harald J. van Heerde (2018), "Delusion in Attribution: Caveats in Using Attribution for Multimedia Budget Allocation," *Journal of Marketing Research*, 55 (5), 667–85.

Datta, Hannes, George Knox, and Bart J. Bronnenberg (2017), "Changing Their Tune: How Consumers' Adoption of Online Streaming Affects Music Consumption and Discovery," *Marketing Science*, 37 (1), 5–21.

Dean, Sarah, Sarah Rich, and Benjamin Recht (2020), "Recommendations and User Agency: The Reachability of Collaboratively-Filtered Information," in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. New York: Association for Computing Machinery, 436–45.

Deng, Yiting and Carl F. Mela (2018), "TV Viewing and Advertising Targeting," *Journal of Marketing Research*, 55 (1), 99–118.

Dwork, Cynthia, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel (2012), "Fairness Through Awareness," in *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*. New York: Association for Computing Machinery, 214–26.

Edelman, Benjamin, Michae Ostrovsky, and Michael Schwarz (2007), "Internet Advertising and the Generalized Second-Price Auction: Selling Billions of Dollars Worth of Keywords," *American Economic Review*, 97 (1), 242–59.

eMarketer (2020a), "Smartphone Users and Penetration in Worldwide," (accessed January 18, 2022), https://forecasts-na1.emarketer.com/5a4fa955d8690c0c28d1f3cc/5a4fa7c2d8690c0c28d1f3b8.

eMarketer (2020b), "US Total Media Ad Spending, by Media," (accessed January 18, 2022), https://forecasts-na1.emarketer.com/584b26021403070290f93a2f/5851918b0626310a2c186b4c.

eMarketer (2019), "Time Spent with Mobile, US," (accessed January 18, 2022), https://forecasts-na1.emarketer.com/584b26021403070290f93a96/5851918b0626310a2c186b39.

Godes, David and Dina Mayzlin (2004), "Using Online Conversations to Study Word-Of-Mouth Communication," *Marketing Science*, 23 (4), 545–60.

Han, Suk Won and René Marois (2014), "Functional Fractionation of the Stimulus-Driven Attention Network," *Journal of Neuroscience*, 34 (20), 6958–69.

He, Xinran, Junfeng Pan, Ou Jin, Tianbing Xu, Bo Liu, Tao Xu, et al (2014), "Practical Lessons from Predicting Clicks on Ads at Facebook," in *Proceedings of the Eighth International Workshop on Data Mining for Online Advertising*. New York: Association for Computing Machinery, 1–9.

Helson, Harry (1948), "Adaptation-Level as a Basis for a Quantitative Theory of Frames of Reference," *Psychological Review*, 55 (6), 297–313.

Hirano, Keisuke, Guido W. Imbens, and Geert Ridder (2003), "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score," *Econometrica*, 71 (4), 1161–89.

Hoch, Stephen J., Eric T. Bradlow, and Brian Wansink (1999), "The Variety of an Assortment," *Marketing Science*, 18 (4), 527–46.

Jeziorski, Przemyslaw and Ilya Segal (2015), "What Makes Them Click: Empirical Analysis of Consumer Demand for Search Advertising," *American Economic Journal: Microeconomics*, 7 (3), 24–53.

Kahneman, Daniel (1973), *Attention and Effort*. Englewood Cliffs, NJ: Prentice Hall.

Kim, Jaehwan, Greg M. Allenby, and Peter E. Rossi (2002), "Modeling Consumer Demand for Variety," *Marketing Science*, 21 (3), 229–50.

Lattimore, Tor and Csaba Szepesvári (2020), *Bandit Algorithms*. Cambridge, UK: Cambridge University Press.

Li, Hongshuang (Alice) and P.K. Kannan (2014), "Attributing Conversions in a Multichannel Online Marketing Environment: An Empirical Model and a Field Experiment," *Journal of Marketing Research*, 51 (1), 40–56.

Ling, Xiaoliang, Weiwei Deng, Chen Gu, Hucheng Zhou, Cui Li, and Feng Sun (2017), "Model Ensemble for Click Prediction in Bing Search Ads," in *Proceedings of the 26th International Conference on World Wide Web Companion*. New York: Association for Computing Machinery, 689–98.

Little, R. and L. Yau (1996), "Intent-to-Treat Analysis for Longitudinal Studies with Drop-Outs," *Biometrics*, 52 (4), 1324–33.

McAlister, Leigh (1982), "A Dynamic Attribute Satiation Model of Variety-Seeking Behavior," *Journal of Consumer Research*, 9 (2), 141–50.

McGranaghan, Matthew, Jura Liaukonyte, and Kenneth C. Wilbur (2021), "TV Ad Viewability: How Viewer Tuning, Presence and Attention Respond to Ad Content," working paper, https://kennethcwilbur.github.io/website/MLW2021%20TV%20Ad%20Viewability.pdf.

McMahan, H. Brendan, Gary Holt, D. Sculley, Michael Young, Dietmar Ebner, Julian Grady, et al (2013), "Ad Click Prediction: A View from the Trenches," in *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: Association for Computing Machinery, 1222–30.

Mirrokni, Vahab, S. Muthukrishnan, and Uri Nadav (2010), "Quasi-Proportional Mechanisms: Prior-Free Revenue Maximization," in *Latin American Symposium on Theoretical Informatics*. New York: Springer, 565–76.

Naik, Prasad A., Murali K. Mantrala, and Alan G. Sawyer (1998), "Planning Media Schedules in the Presence of Dynamic Advertising Quality," *Marketing Science*, 17 (3), 214–35.

Nair, Harikesh S., Sanjog Misra, William J. Hornbuckle IV, Ranjan Mishra, and Anand Acharya (2017), "Big Data and Marketing Analytics in Gaming: Combining Empirical Models and Field Experimentation," *Marketing Science*, 36 (5), 699–725.

Pieters, Rik, Edward Rosbergen, and Michel Wedel (1999), "Visual Attention to Repeated Print Advertising: A Test of Scanpath Theory," *Journal of Marketing Research*, 36 (4), 424–38.

Rafieian, Omid (2020), "Revenue-Optimal Dynamic Auctions for Adaptive Ad Sequencing," working paper, https://omidraf.github.io/data/AdSeqAuction.pdf.

Rafieian, Omar (2022), "Optimizing User Engagement Through Adaptive Ad Sequencing." working paper.

Rafieian, Omid and Hema Yoganarasimhan (2021), "Targeting and Privacy in Mobile Advertising," *Marketing Science*, 40 (2), 193–218.

Ratner, Rebecca K., Barbara E. Kahn, and Daniel Kahneman (1999), "Choosing Less Preferred Experiences for the Sake of Variety," *Journal of Consumer Research*, 26 (1), 1–15.

Redden, Joseph P. (2008), "Reducing Satiation: The Role of Categorization Level," *Journal of Consumer Research*, 34 (5), 624–34.

Rosenbaum, Paul R. and Donald B. Rubin (1983), "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70 (1), 41–55.

Rutz, Oliver J. and Randolph. E. Bucklin (2011), "From Generic to Branded: A Model of Spillover in Paid Search Advertising," *Journal of Marketing Research*, 48 (1), 87–102.

Sahni, Navdeep S. (2015), "Effect of Temporal Spacing Between Advertising Exposures: Evidence from Online Field Experiments," *Quantitative Marketing and Economics*, 13 (3), 203–47.

Sahni, Navdeep S. (2016), "Advertising Spillovers: Evidence from Online Field Experiments and Implications for Returns on Advertising," *Journal of Marketing Research*, 53 (4), 459–78.

Schumann, David W., Richard E. Petty, and D. Scott Clemons (1990), "Predicting the Effectiveness of Different Strategies of Advertising Variation: A Test of the Repetition-Variation Hypotheses," *Journal of Consumer Research*, 17 (2), 192–202.

Tellis, Gerard J. (2003), *Effective Advertising: Understanding When, How, and Why Advertising Works*. Thousand Oaks, CA: SAGE Publications.

Theocharous, Georgios, Philip S. Thomas, and Mohammad Ghavamzadeh (2015), "Personalized Ad Recommendation Systems for Life-Time Value Optimization with Guarantees," in *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence*. New York: Association for Computing Machinery, 1806–12.

Tuchman, Anna E., Harikesh S. Nair, and Pedro M. Gardete (2018), "Television Ad-Skipping, Consumption Complementarities and the Consumer Demand for Advertising," *Quantitative Marketing and Economics*, 16 (2), 111–74.

Tunuguntla, Srinivas and Paul R. Hoban (2021), "A Near-Optimal Bidding Strategy for Real-Time Display Advertising Auctions," *Journal of Marketing Research*, 58 (1), 1–21.

Varian, Hal R. (2007), "Position Auctions," *International Journal of Industrial Organization*, 25 (6), 1163–78.

Wilbur, Kenneth C. (2008), "A Two-Sided Empirical Model of Television Advertising and Viewing Markets," *Marketing Science*, 27 (3), 356–78.

Wilbur, Kenneth C. (2016), "Advertising Content and Television Advertising Avoidance," *Journal of Media Economics*, 29 (2), 51–72.

Wilbur, Kenneth C., Linli Xu, and David Kempe (2013), "Correcting Audience Externalities in Television Advertising," *Marketing Science*, 32 (6), 892–912.

Yao, Song and Carl F. Mela (2011), "A Dynamic Model of Sponsored Search Advertising," *Marketing Science*, 30 (3), 447–68.

Yoganarasimhan, Hema (2013), "The Value of Reputation in an Online Freelance Marketplace," *Marketing Science*, 32 (6), 860–91.

Zantedeschi, Daniel, Eleanor M. Feit, and Eric T. Bradlow (2017), "Measuring Multichannel Advertising Response," *Management Science*, 63 (8), 2706–28.