# Web Appendix for Variety Effects in Mobile Advertising

Omid Rafieian* Cornell University

Hema Yoganarasimhan* University of Washington

October 10, 2021

## Contents

### Web Appendix



*These materials have been supplied by the authors to aid in the understanding of their paper. The AMA is sharing these materials at the request of the authors. Please address all correspondence to: or83@cornell.edu, hemay@uw.edu.

# A Web Appendix A: Details of Sampling Procedure

We sample a set of sessions from the full data as follows:

- *Users:* To identify users with untruncated history, we split the data into two parts and make two sets of users: (1) $\mathcal{U}_1$ that consists of users generated at least one impression from 30 September 2015 to 21 October 2015, and (2) $\mathcal{U}_2$ that consists of users who generated at least one impression from 22 October 2015 to 30 October 2015. We then sample all users who are available in the second set but not in the first set, i.e., $\mathcal{U}_2 \setminus \mathcal{U}_1$. The fact that the user was not available in the first set suggests that we have the entire observed activity for that user.
- *App:* Within the set of users who satisfy the above condition, we then exclusively focus on their impressions in the most popular app. This is a messaging app that is widely used in the country, and generates over 30% of the total traffic observed in the ad-network. Just focusing on the top app allows us to hold the context of the app constant and cleanly derive the causal effect of the variety of previous ads.

  The empirical CDF of the session length in our sample is shown in Figure W1.

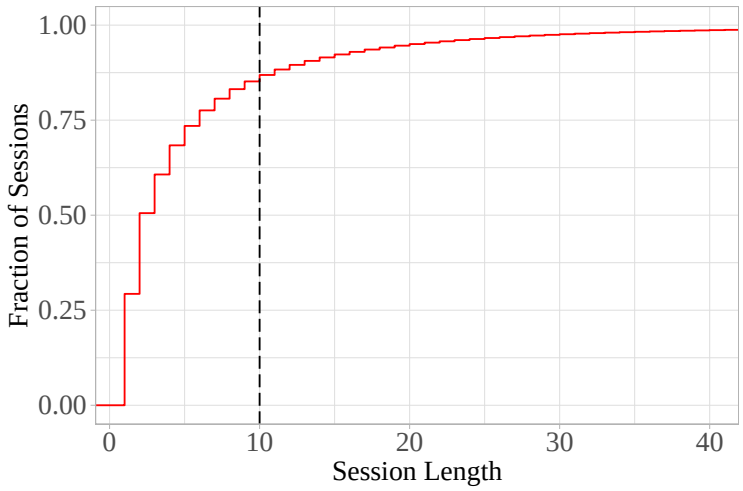


Figure W1: Session length (truncated at 40).

# B Web Appendix B: Alternative Specifications for Preliminary Analysis

## B.1 Shannon Entropy as an Alternative Measure of Variety

We establish the results presented in Table 2 of the paper for a different measure of variety. In particular, we consider Shannon entropy, which is a widely used metric for dispersion and diversity in the information theory literature (Shannon, 1948). In the marketing literature, this measure has been used to quantify dispersion or variety of variables (Godes and Mayzlin, 2004). For a sequence $\langle A_{i,t} \rangle_{t=1}^{T_i}$, let $I_{i,a,t}$ denote the number of times ad $a$ has been shown in session $i$ prior to exposure $t$. We can define Shannon entropy as follows:

$$\textit{Shannon}_{i,t} = -\sum_{a|I_{i,a,t}>0} \frac{I_{i,a,t}}{t-1} \log\left(\frac{I_{i,a,t}}{t-1}\right). \tag{W1}$$

Intuitively, this measure captures the *amount of information* in the past sequence. More practically, it translates into number of bits required to store the sequence. As such, it takes a higher value when the variety of prior ads is higher. We replace $V_{i,t}$ with $\textit{Shannon}_{i,t}$ and estimate the models in Table 2. We present the results of these models in Table W1. As shown in this table, all qualitative findings in Table 2 remain the same in our new analysis.

| | *Dependent variable: Click ($Y_{i,t}$)* | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| $\textit{Shannon}_{i,t}$ | 0.00715*** | 0.00323*** | 0.00323*** | 0.00269*** |
| | (26.35) | (11.22) | (11.20) | (8.48) |
| $\textit{Freq}_{i,t}$ | | | | −0.00035** |
| | | | | (-2.55) |
| $\textit{Space}_{i,t}$ | | | | 0.00035*** |
| | | | | (5.31) |
| Ad FE | ✓ | ✓ | ✓ | ✓ |
| Exposure Number FE | ✓ | ✓ | ✓ | ✓ |
| Targeting Variables FE | | ✓ | ✓ | ✓ |
| Session Length FE | | | ✓ | ✓ |
| $R^2$ | 0.007 | 0.009 | 0.009 | 0.009 |
| Adjusted $R^2$ | 0.007 | 0.008 | 0.009 | 0.009 |
| No. of Obs. | 1,993,542 | 1,993,542 | 1,993,542 | 1,993,542 |
| *Note:* | | | | *p<0.05; **p<0.01; ***p<0.001 |

Table W1: Preliminary results on the effect of variety on click outcome when using Shannon entropy to measure variety.

## B.2 Preliminary Results with Logistic Regression

Since our outcome variable is binary, we replicate the results of Table 2 with a logistic regression model. We present the estimates of our logistic regression model in Table W2. The results show the same patterns as our main analysis. Please notice that the discrepancy in the number of observations

between Table 2 and Table W2 is due to the fact that the logistic regression categories for which the outcome has no variation (e.g., no click for a specific ad).

| | *Dependent variable: Click ($Y_{i,t}$)* | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Variety ($V_{i,t}$) | 0.15265*** | 0.06922*** | 0.06922*** | 0.05016*** |
| | (0.00569) | (0.00608) | (0.00608) | (0.00642) |
| $Freq_{i,t}$ | | | | −0.05381** |
| | | | | (0.00914) |
| $Space_{i,t}$ | | | | 0.01372*** |
| | | | | (0.00433) |
| Ad FE | ✓ | ✓ | ✓ | ✓ |
| Exposure Number FE | ✓ | ✓ | ✓ | ✓ |
| Targeting Variables FE | | ✓ | ✓ | ✓ |
| Session Length FE | | | ✓ | ✓ |
| No. of Obs. | 1,986,048 | 1,986,048 | 1,989,602 | 1,989,602 |
| *Note:* | | | | *p<0.05; **p<0.01; ***p<0.001 |

Table W2: Preliminary results on the effect of breadth of variety on click outcome using a logistic regression model. Robust standard errors are presented in paranthesis.

# C Web Appendix C: Proofs

## C.1 Proof of Proposition 1

The distribution of propensity scores for ads is fully determined by the allocation rule in a quasi-proportional auction. That is, for any ad $a$ competing for the impression $t$ in session $i$, the probability that this ad wins an impression is denoted by $\pi_{i,t}(a)$, characterized as follows:

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

where $\mathcal{C}_{i,t}$ is the set of ads competing in the auction for exposure $t$ in session $i$, and $b_{i,a}m_{i,a}$ is ad $a$'s quality-adjusted bid which is the product of the ad's bid ($b_{i,a}$) and quality score ($m_{i,a}$). If we observe all the components of this equation, the distribution of propensity scores for ads is a function of observed variables by default. However, we do not observe the quality scores in our data. So we need to show that the distribution of propensity scores is still fully identified even without observing the quality scores. We first use the following lemma that helps us establish the identifiability of the distribution of propensity scores within a specific sub-sample of our sessions:

**Lemma 1.** *Let $G$ denote a set of sessions where the auction is identical for any two sessions $i$ and $j$ that $i, j \in G$. The distribution of propensity scores for ads is identified if we observe the actual ad assignment in our data.*

*Proof.* If we know the actual ad assignments for impressions in $G$, the proportion of impressions in $G$ that show $a$ in is an accurate estimate of the propensity score for that ad because the distribution of ad assignment is identical across impressions in $G$. □

In light of this lemma, if we know the actual ad assignments in the data, we do not need to observe bids or quality scores to identify the distribution of propensity scores in any group $G$ where the auction is identical across impressions within that group. Now, if we show that such partitioning or stratification of our sessions is feasible, our proof is complete. We use the fact that the ad-network does not update quality scores throughout the day. Hence, to make sure that the quality scores remain constant in all the impressions within a partition, it is sufficient to partition our sessions such that sessions in different days are not in the same partition. Finally, since we directly observe bids, we only need to show that we can identify $\mathcal{C}_{i,t}$. This set can only vary if sessions are different in two dimensions: (1) targeting characteristics, because some ads may decide to exclude some sessions based on their targeting characteristics, and/or (2) time because some ad campaigns may be unavailable at some points in time. Since we observe all the targeting characteristics as well as the exact timestamp of each impression, we can identify the groups where all the sessions have the same $\mathcal{C}_{i,t}$, and this completes our proof.

## C.2 Proof of Proposition 2

We only need to show that if two exposures share the same targeting characteristics and happen the same timestamp $\tau$, their auctions will be identical. Let $i$ and $j$ denote two sessions, and $t$ and $t'$ refer to the corresponding exposure numbers at timestamp $\tau$, respectively, in these two sessions. For

the two auctions for these impressions to be identical, we need to satisfy the following condition:

$$\forall a, \quad \mathbb{1}(a \in \mathcal{C}_{i,t}) \frac{b_{i,a} m_{i,a}}{\sum_{k \in \mathcal{C}_{i,t}} b_{i,k} m_{i,k}} = \mathbb{1}(a \in \mathcal{C}_{j,t'}) \frac{b_{j,a} m_{j,a}}{\sum_{k \in \mathcal{C}_{j,t'}} b_{j,k} m_{j,k}}. \tag{W2}$$

We now show that this equality holds if these two sessions share the same targeting characteristics. This is because all the elements of the LHS and RHS become identical under this condition. We establish this in more detail below:

1. *Equality in bids:* In our setting, ads are only allowed to submit a single bid at any given point in time. Therefore, an advertiser's bid can be different across sessions (i.e., $b_{i,a} \neq b_{j,a}$) if and only if the advertiser changed his bid between the two sessions. Further, if an advertiser changes his bid, it becomes effective only in the next hour. Thus, if $i$ and $j$ are sessions with the same targeting characteristics and impressions $(i, t)$ and $(j, t')$ happen around the same timestamp $\tau$ (as broad as an hour), then for all $a$, $b_{i,a} = b_{j,a}$.

2. *Equality in quality scores:* A unique feature of our setting is that unlike most platforms, this platform does not customize quality scores and only uses an aggregate measure for every ad. Every few days, the platform updates these quality scores. Thus, for sessions $i$ and $j$ in the same day, we have $m_{i,a} = m_{j,a}$. This implies that for two impressions $(i, t)$ and $(j, t')$ with the same targeting characteristics and around the same timestamp $\tau$ (as broad as an hour), the we have $m_{i,a} = m_{j,a}$ for all $a$.

3. *Equality in the set of participants:* If there exists an ad that participates in auction for session $i$ but not session $j$ at timestamp $\tau$ (or vice versa), then Equation (W2) is violated. To show that this is not the case for sessions with the same targeting characteristics, we first discuss potential sources of discrepancy in the set of participants and then show that these sources are blocked when $i$ and $j$ share the same targeting characteristics:

   (a) *Difference in targeting:* Advertisers can target their ads based on app category, province, connectivity type, time of the day, MSP, and smartphone brand. As such, each session has a set of targeting characteristics. Hence, we may have $a \notin \mathcal{C}_{i,t}$ and $a \in \mathcal{C}_{j,t'}$ because ad $a$ decided to target session $i$ but not session $j$ (or vice versa). For example, if an advertiser selects Huawei and LG as the set of smartphone brand categories he wants to target, his ad will not be shown to any Samsung users, because Samsung is excluded from her targeting set. Now, if the smartphone brand is Huawei in session $i$ and Samsung in session $j$, then this ad will be in $\mathcal{C}_{i,t}$ but not $\mathcal{C}_{j,t'}$. However, this source is fully blocked if $i$ and $j$ have the same targeting characteristics. Thus, for any $a \in \mathcal{C}_{i,t}$, we have $a \in \mathcal{C}_{j,t'}$.

   (b) *Difference in availability over time:* An advertiser's campaign availability over time can change due to three possible reasons: entry, exit, and budget. Figure W2 illustrates this point by showing an ad's availability during the time of study. As such, if impression $(i, t)$ occurs during a time when ad $a$ is unavailable (due to entry, exit, or budget) and impression $(j, t')$ occurs during a time when $a$ is available, then we have $a \notin \mathcal{C}_{i,t}$ but

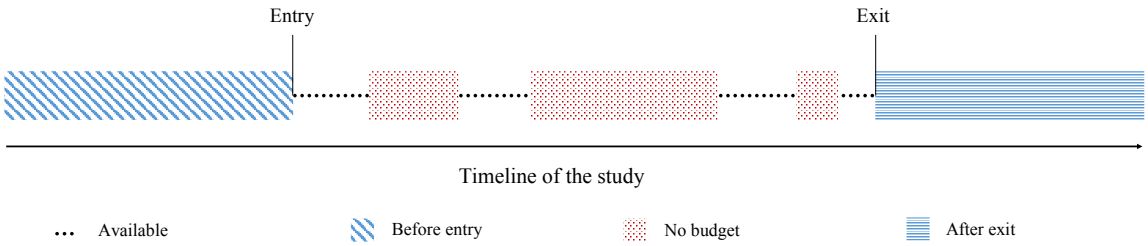


Figure W2: Availability of an ad in the timeline of the study due to entry, exit, and budget exhaustion

$a \in \mathcal{C}_{j,t'}$. However, this source is also blocked if $(i,t)$ and $(j,t')$ happen at the same timestamp $\tau$ or in its local neighborhood (empirically, even a gap of an hour does not induce much change in the auctions).

Thus, for impressions $(i,t)$ and $(j,t')$ around the same timestamp with the same targeting characteristics, we have $\mathcal{C}_{i,t} \equiv \mathcal{C}_{j,t'}$.

Together, for impressions $(i,t)$ and $(j,t')$ around the same timestamp with the same targeting characteristics, we have the equality in Equation (W2), and this completes the proof.

# D Web Appendix D for Propensity Scores

In this section, we first describe how we estimate the propensity scores for the treatment variable and then show how we assess covariate balance for pre-treatment variables.

## D.1 Estimation of Propensity Scores

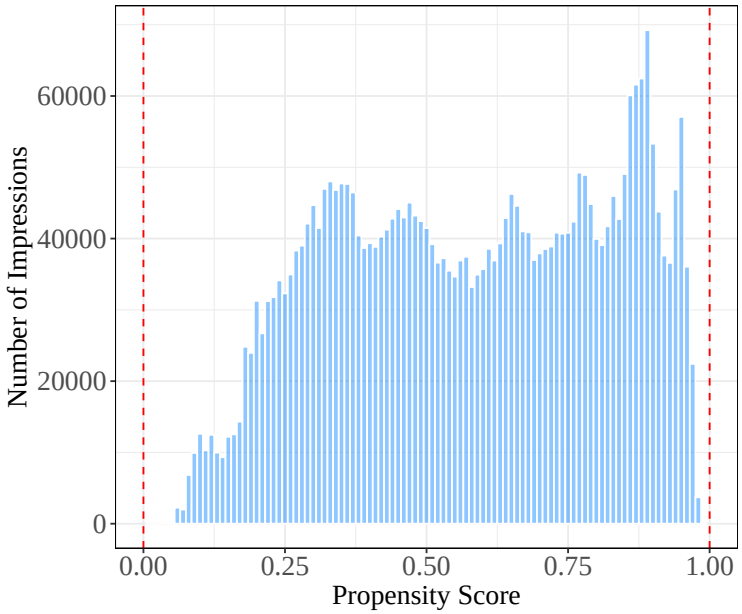


Figure W3: Histogram of estimated propensity scores

We now describe the procedure to estimate the propensity scores of the treatment. As mentioned in the main text of the paper, we can use any learner for this purpose since it is essentially a prediction task. We use XGBoost, which is a fast and scalable version of Gradient Boosting machines that have been used to estimate propensity scores in the past literature (McCaffrey et al., 2013). Like any supervised learning algorithm, XGBoost requires an outcome variable and a set of covariates as inputs for training. The outcome variable or label for this task is the treatment assignment observed in the data ($W_{i,t}$). The set of covariates needed to accurately estimate propensity scores are:

- All the targeting variables $X_{i,t-1}$: province, hour of the day, smartphone brand, MSP, and connectivity type. We include dummy variables for each of these variables. Please notice that the subscript $t-1$ is just for notational consistency, since all these variables largely remain the same for exposure $t$ as well.
- The exact timestamp of the impression to capture any change in the auction, including entry and exit of ads at different points of time.
- The exact latitude and longitude of the user. This is an unnecessary but harmless control.
- A dummy for each ad that indicates whether that particular ad has been shown in the first $t-2$ exposures of the current session. For example, suppose that the sequence of ads shown from exposure 1 to $t-2$ is $\langle A, B, A, C, D\rangle$. Our covariates for all the distinct ads shown $A$, $B$, $C$, and $D$ take value one, whereas for the rest of ads these dummy variables are zero. This set of dummy variables helps incorporate the fact that $e(W_{i,t}) = \sum_{a \notin H_{i,t-2}} \pi_{i,t-1}(a)$.

We use this set of covariates to estimate the propensity scores. To avoid overfitting, we use an early stopping criterion that stops our XGBoost model once two iterations give the same logarithmic loss.

We obtain our propensity estimates and plot their histogram in Figure W3. We see extensive variation in the estimated propensities. Notably, there is no deterministic propensity score, i.e., $0 < \hat{e}(W_{i,t}) < 1$. This ensures that we have the overlap assumption necessary for causal inference.

### D.2 Covariate Balance

For covariate balance, we need to ensure that the IPW-adjusted distribution of pre-treatment variables is the same across control and treatment groups. To assess covariate balance, we use the standardized bias measure, which is the commonly used in the literature (McCaffrey et al., 2013). For any pre-treatment variable $X$, let $\bar{X}_{W=1}$ denote the population mean of variable $X$ when assigned to the treatment. We denote the IPW-adjusted mean of this variable by $\bar{X}_{W=1}^{\textit{IPW-adjusted}}$ and characterize it as:

$$\bar{X}_{W=1}^{\textit{IPW-adjusted}} = \frac{\sum_{j=1}^{N} \frac{\mathbb{1}(W_j=1)}{\hat{e}(W_j)} X_j}{\sum_{j=1}^{N} \frac{\mathbb{1}(W_j=1)}{\hat{e}(W_j)}}, \tag{W3}$$

where $j$ is the subscript for each impression in our data and $N$ denotes the total number of impressions in the sample. Using the definition of $\bar{X}_{W=1}^{\textit{IPW-adjusted}}$, for any variable $X$, we define the following standardized bias (SB) measure:

$$\textit{SB}(X) = \frac{|\bar{X}_{W=1}^{\textit{IPW-adjusted}} - \bar{X}|}{\sigma_X}, \tag{W4}$$

where $\bar{X}$ is the population mean for variable $X$, and $\sigma_X$ denotes its standard deviation. In general, we need to specify a threshold $\alpha$ for the standardized bias such that if $\textit{SB}(X) < \alpha$, we can conclude balance for variable $X$. The conventional norm in the literature is 0.2 or sometimes 0.1 (McCaffrey et al., 2013; Austin, 2009). We take a more conservative measure and assess balance only if $\textit{SB}(X) < 0.02$.

We check the balance for all our targeting subcategories. Before adjusting for propensity scores, 25 subcategories are unbalanced, i.e., standardized bias is greater than 0.02. As expected, 21 of these subcategories are provinces, and 4 of them are hours of the day. We do not observe any covariate imbalance for subcategories within smartphone brand, MSP, or connectivity type, since these variables were not being used for targeting. After adjusting for propensity scores according to Equation (W3), we achieve balance for all the targeting subcategories.

Next, we check balance for the past variety $V_{i,t-1}$. Again, in the unadjusted case, we are unable to assess covariate balance. However, after adjusting for propensity weights, we assess covariate balance. We further check the balance for the dummy variables indicating whether or not each ad has been shown before in the session. While we have covariate imbalance before adjusting for propensity scores for some ads, we achieve balance for all the ads after adjusting for propensity weights.

# E Web Appendix E: Imputation Procedure for Dynamic Selection

We now present the step-by-step imputation algorithm. Let $\tilde{D}$ be the complementary data that we use for imputation.

- *Step 1:* For any session $i$ that ended in $T_i$ exposures, we can impute the timestamp for the exposure $T_i + 1$ that would have happened had the user stayed. We denote this timestamp for the imputed impression as $\tau^*_{i,T_i+1}$. Since each exposure lasts one minute, we add 60 seconds to $\tau_{i,T_i}$ to obtain $\tau^*_{i,T_i+1}$.
- *Step 2:* For any timestamp $\tau^*_{i,T_i+1}$, we find an impression $j$ in the complementary data ($j \in \tilde{D}$) with the same targeting characteristics as session $i$ at timestamp $\tau^*_{i,T_i+1}$.[1] Let $\tilde{A}_j(\tau^*_{i,T_i+1})$ denote the ad shown in impression $j$ at timestamp $\tau^*_{i,T_i+1}$.
- *Step 3:* We impute the ad that would have been shown in session $i$ at exposure $T_i + 1$ with the ad found in the complementary data since it represents the ad that could have been shown in the $(T_i + 1)^{\text{th}}$ exposure of session $i$ had it not ended. Hence, $A^*_{i,T_i+1} = \tilde{A}_j(\tau^*_{i,T_i+1}) \sim \mathcal{A}_j(\tau^*_{i,T_i+1})$.

[1] If there are more than one candidate, we randomly choose one impression. Likewise, if there is no impression available at the exact second, we choose the impression that is temporally the closest to $\tau^*_{i,T_i+1}$.

# F Web Appendix F: Robustness Checks for the Main Effects

## F.1 Robustness Checks with Logistic Regression

Our main results in Table 3 of the paper uses a linear probability model. There are several reasons why we focused on linear probability models instead of nonlinear models such as logistic regression. First, we are interested in partial effects for a model where independent variables have a very restricted in the values they can take. In fact, all the covariates are binary variables in our main specifications. Issues with the linear probability models generally arise when independent variables can take extreme values (Wooldridge, 2010). Second, our goal is to estimate partial effects not prediction. As such, we need not worry about the probabilities not lying within the [0,1] range. For the goal of estimating partial effects, we can show that OLS produces consistent and unbiased estimates of coefficients of the linear specification (Wooldridge, 2010). Third, linear probability model flexibly allows for including a very conservative set of covariates and fixed effects while using the full data, which makes it computationally advantageous over nonlinear models such as logistic regression.

However, to check the robustness of our main results, we replicate Table 3 using logistic regression. We present the results of this logistic regression in Table W3. As shown in this table, the results of our logistic regression model generate the same qualitative and even quantitative results.

| | *Dependent variable: Click ($Y^*_{i,t}$)* | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Treatment ($W_{i,t}$) | 0.11808*** | 0.13371*** | 0.12839*** | 0.14175*** |
| | (0.01540) | (0.01349) | (0.01547) | (0.01505) |
| IPW Adjustment | ✓ | | ✓ | ✓ |
| Imputed Sample | ✓ | ✓ | | ✓ |
| Exposure ($t$) FE | ✓ | ✓ | ✓ | ✓ |
| $Freq_{i,t}$ Indicators | ✓ | ✓ | ✓ | |
| $Space_{i,t}$ Indicators | ✓ | ✓ | ✓ | |
| $A^*_{i,t}$ FE | ✓ | ✓ | ✓ | ✓ |
| $A_{i,t-1}$ FE | ✓ | ✓ | ✓ | ✓ |
| No. of Obs. | 2,395,948 | 2,395,948 | 1,987,418 | 2,395,948 |
| *Note:* | | | | *p<0.05; **p<0.01; ***p<0.001 |

Table W3: Average effects of the variety treatment on the CTR using logistic regression. Numbers reported in parentheses are robust standard errors.

### F.2 Robustness Checks with Additional Pre- and Post-Treatment Controls

Our main specification takes both pre- and post-treatment confounding into account. Nevertheless, in a series of robustness checks, we now consider even more conservative models with extra controls for pre- and post-treatment variables. One downside of these models is that they ignore much of the exogenous variation in the data and may lack sufficient power. However, they present a conservative test of our main effects – if we are able to find that the effects of variety continue to be directionally correct and significant in these models, it gives us even more confidence in our findings. We now describe these models and present their results in Table W4.

1. **Controlling for targeting variables and past variety:** In the main specification, we control for pre-treatment confounding using IPWs, where the propensity scores are of being assigned to the treatment ($W_{i,t}$) are estimated as a function of all the targeting variables as well as the prior set of ads shown in the session.

   In this model, in addition to the propensity score correction, we also include the covariates that determine the propensity scores directly into the model specification. To do so, we first define *Target* as the interaction of all targeting variables and hour/day. That is, two impressions with the same value of *Target* share the same province, smartphone brand, MSP, connectivity type, and happen at the same hour on the same day. In light of Proposition 2, we know that impressions within a specific targeting area have the same ad allocation distribution. Overall, it gives us 90,074 distinct values for *Target*. Next, as controls for the past set of ads shown in the session, we also include the indicators for different levels of $V_{i,t-1}$. The overall specification of this model is as follows:

$$\begin{aligned} Y^*_{i,t} =& \beta W_{i,t} + \sum_q \gamma_q \mathbb{1}(\textit{Freq}_{i,t} = q) + \sum_s \delta_s \mathbb{1}(\textit{Space}_{i,t} = s) + \sum_r \theta_r \mathbb{1}(V_{i,t-1} = r) \\ &+ \alpha_0(A^*_{i,t}) + \alpha_1(A_{i,t-1}) + \zeta_t + \eta_{\textit{Target}} + \epsilon_{i,t}, \end{aligned} \tag{W5}$$

   where $\eta_{\textit{Target}}$ captures the fixed effects for *Target*, and $\theta_r$ is the coefficient for level $r$ of past variety. We present the results of this model in the first column of Table W4. The significant and positive coefficient for the treatment effects confirms our main effect.

2. **Controlling for user fixed effects and targeting variables:** In this model, we include a different set of controls for the discrepancy in the treatment assignment – user fixed effects and hour/day fixed effects. Controlling for user fixed effects ensures that we only consider the variation within users, so by default it is robust to the selection caused by the differences between users (e.g., users who are more likely to click may be assigned to the treatment condition more often). These fixed effects can also capture the targeting variables that are likely to be fixed within the user (e.g., smartphone brand) and the hour/time of the day (e.g., set of ads targeting that time of day). As before, we also control for the past set of ads shown,

where we use different levels of $V_{i,t-1}$. As such, our model is:

$$\begin{aligned} Y^*_{i,t} =& \beta W_{i,t} + \sum_q \gamma_q \mathbb{1}(\textit{Freq}_{i,t} = q) + \sum_s \delta_s \mathbb{1}(\textit{Space}_{i,t} = s) + \sum_r \theta_r \mathbb{1}(V_{i,t-1} = r) \\ &+ \alpha_0(A^*_{i,t}) + \alpha_1(A_{i,t-1}) + \zeta_t + \eta_{\textit{User}} + \kappa_{\textit{Hour-Day}} + \epsilon_{i,t}, \end{aligned} \tag{W6}$$

where $\eta_{\textit{User}}$ and $\kappa_{\textit{Hour-Day}}$ are fixed effects for users and hour-day combination. Overall, this model comes with 71,945 separate user fixed effects and 217 separate hour-day fixed effects. As before, we use an IPW-adjusted regression to estimate this model and present the results in the second column of Table W4. The results show the same pattern as before: an increase in ad variety leads to higher CTR on the next ad.

3. **Controlling for session fixed effects:** Next, we go one step further and make the model even more restrictive by including the session-level fixed effects. It clearly controls for the ad allocation distribution, since the auction is the same for all the seven exposures in a session that we focus on. This gives us 583,694 distinct categories to control for. We use an IPW-adjusted regression to estimate the following model:

$$\begin{aligned} Y^*_{i,t} =& \beta W_{i,t} + \sum_q \gamma_q \mathbb{1}(\textit{Freq}_{i,t} = q) + \sum_s \delta_s \mathbb{1}(\textit{Space}_{i,t} = s) + \sum_r \theta_r \mathbb{1}(V_{i,t-1} = r) \\ &+ \alpha_0(A^*_{i,t}) + \alpha_1(A_{i,t-1}) + \zeta_t + \eta_{\textit{Session}} + \epsilon_{i,t}, \end{aligned} \tag{W7}$$

where $\eta_{\textit{Session}}$ controls for session fixed effects. We present the results of this model in the third column of Table W4. Once again, we find that our main effects are robust, even when we use such a narrow lens on our comparison, and a specification that soaks up the exogenous variation across sessions.

4. **Controlling for interaction of all post-treatment variables:** Now, we consider a model with additional post-treatment controls. As discussed before, our approach to address post-treatment confounding is based on a *ceteris paribus* interpretation. Therefore, we now include controls that capture more complex forms of post-treatment confounding. We start by defining a variable $\textit{Current}_{i,t}$ that captures the collective information about the current ad as the interaction of $A^*_{i,t}$, $\textit{Freq}_{i,t}$, and $\textit{Space}_{i,t}$. That is, we make a separate category for each unique combination of these three variables. This gives us 7,174 categories of $\textit{Current}_{i,t}$ that we need to control for. We can write this model as follows:

$$Y^*_{i,t} = \beta W_{i,t} + \alpha_1(A_{i,t-1}) + \zeta_t + \eta_{\textit{Current}} + \epsilon_{i,t}, \tag{W8}$$

where $\eta_{\textit{Current}}$ accounts for the fixed effects of each value of $\textit{Current}_{i,t}$. Note that the above model does not have separate controls for $A^*_{i,t}$, $\textit{Freq}_{i,t}$, and $\textit{Space}_{i,t}$, but instead controls for $\textit{Current}_{i,t}$. We again use an IPW-adjusted regression to estimate our main results. The results for this model are shown in the fourth column of Table W4. Again, we see that the treatment effect is significant and positive, which shows the robustness of our results to more restrictive post-treatment controls.

| | *Dependent variable: Click ($Y^*_{i,t}$)* | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Treatment ($W_{i,t}$) | 0.00100*** | 0.00121*** | 0.00061** | 0.00201*** | 0.00141*** |
| | (4.70) | (5.79) | (2.42) | (9.01) | (5.99) |
| IPW Adjustment | ✓ | ✓ | ✓ | ✓ | ✓ |
| Exposure ($t$) FE | ✓ | ✓ | ✓ | ✓ | ✓ |
| $\mathit{Freq}_{i,t}$ Indicators | ✓ | ✓ | ✓ | | |
| $\mathit{Space}_{i,t}$ Indicators | ✓ | ✓ | ✓ | | |
| $A^*_{i,t}$ FE | ✓ | ✓ | ✓ | | |
| $A_{i,t-1}$ FE | ✓ | ✓ | ✓ | ✓ | |
| $\mathit{Target}_{i,t}$ FE | ✓ | | | | |
| $V_{i,t-1}$ Indicators | ✓ | ✓ | ✓ | | |
| User FE | | ✓ | | | |
| Hour-Day FE | | ✓ | | | |
| Session FE | | | ✓ | | |
| $\mathit{Current}_{i,t}$ FE | | | | ✓ | |
| $\mathit{Post}_{i,t}$ FE | | | | | ✓ |
| No. of Obs. | 2,405,695 | 2,405,695 | 2,405,695 | 2,405,695 | 2,405,695 |
| $R^2$ | 0.063 | 0.075 | 0.257 | 0.009 | 0.0571 |
| Adjusted $R^2$ | 0.026 | 0.046 | 0.018 | 0.006 | 0.0015 |
| *Note:* | | | | | *p<0.05; **p<0.01; ***p<0.001 |

Table W4: Robustness checks on the average effects of the an increase in ad variety on the CTR on the next ad. Numbers reported in parentheses are t-statistics based on robust standard errors.

Next, we go one step further and define a new variable $\mathit{Post}_{i,t}$ by adding another interaction for the previous ad. As such, this variable is the interaction of all four post-treatment controls – $A^*_{i,t}$, $A^*_{i,t}$, $\mathit{Freq}_{i,t}$, and $\mathit{Space}_{i,t}$. The number of distinct categories in $\mathit{Post}_{i,t}$ is 133,845. The model used for this alternative approach is presented below:

$$Y^*_{i,t} = \beta W_{i,t} + \zeta_t + \eta_{\mathit{Post}} + \epsilon_{i,t}, \tag{W9}$$

where $\eta_{\mathit{Post}}$ controls for all the values of the interaction variable $\mathit{Post}_{i,t}$. We use an IPW-adjusted regression to estimate this model and present its results in the fifth column of Table W4. As before, the results show a positive and significant treatment effects.

A few additional points are worth noting here. First, note that all the above models are overly conservative in the sense that they also soak up different sources of exogenous variation depending on their specification. This adversely affects their inferential power. Nevertheless, all of them show a positive and significant effect of variety. Second, in all the models in Table W4, we adjust for IPWs in the regression. If we exclude this adjustment, all the effects become more significant both in terms of the coefficients and t-statistics. Third, we use hourly split of the data to use the results of Proposition 2. It is worth emphasizing that even with shorter splits such as 30 minutes or 10 minutes, the results remain qualitatively the same.

## F.3 Robustness Check with Exact Matching of the Sequence of Ads

| | *Dependent variable: Click ($Y^*_{i,t}$)* | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Treatment ($W_{i,t}$) | 0.00111* | 0.00118* | 0.00121* | 0.00114* |
| | (2.49) | (2.34) | (2.32) | (2.19) |
| IPW Adjustment | ✓ | ✓ | ✓ | ✓ |
| Matching Group FE | ✓ | ✓ | ✓ | ✓ |
| $A_{i,t-1}$ FE | ✓ | ✓ | ✓ | ✓ |
| *Freq*$_{i,t}$ Indicators | | ✓ | | ✓ |
| *Space*$_{i,t}$ Indicators | | ✓ | | ✓ |
| Targeting Subcategories FE | | | ✓ | ✓ |
| No. of Obs. | 2,405,695 | 2,405,695 | 2,405,695 | 2,405,695 |
| $R^2$ | 0.597 | 0.595 | 0.597 | 0.597 |
| Adjusted $R^2$ | 0.026 | 0.026 | 0.027 | 0.027 |
| *Note:* | | | | *p<0.05; **p<0.01; ***p<0.001 |

Table W5: Results from exact matching of the sequence of ads. Numbers reported in parentheses are t-statistics computed based on robust standard errors.

In this analysis, we consider an exact matching model to jointly control for both pre- and post-treatment confounding issues, and fully isolate the treatment effect. As such, our goal is to ensure that within units of a matching group, everything is the same except the treatment. To do so, we match the entire sequence of ads except the ad shown at point $t-1$. Let *Matching*$_{i,t}$ denote the matching group that exposure $t$ in session $i$ belongs to. If there is a session $j$ such that *Matching*$_{j,t}$ = *Matching*$_{i,t}$, then we have $\langle A_{i,1}, \ldots, A_{i,t-2}, A_{i,t}\rangle = \langle A_{j,1}, \ldots, A_{j,t-2}, A_{j,t}\rangle$. Hence, the only difference will be in exposure $t-1$, where the treatment assignment happens. This exact matching procedure ensures that all pre-treatment and post-treatment sequence-related factors are identical across within a group, thereby allowing us to fully isolate the treatment effects. Of the 2,405,695 impressions, we are able to match 968,343 of them, with the total of 82,149 separate matching categories.

We control for the *Matching*$_{i,t}$ category for each observation and run a series of different specifications and present the results in Table W5. Since the propensity can still be different despite being in the same matching group, we use IPW-adjusted regression to estimate our treatment effects under this approach. In the first column, we show the results for a model that regresses the click outcome on ad at point $t$ on the treatment, the specific ad shown at the treatment phase and matching group fixed effects. Although there are 1,410,405 matching categories that substantially reduce our statistical power, we still find a significant and positive coefficient for the treatment effects. Next, for the models in columns 2–4, we add more restrictive controls. Even so, the same patterns emerge.

Notice that the total number of categories is higher than the number of categories that we can match, since the former also includes impressions are left unmatched as category with one

observation. The categories with just one impression in them do not play a role in our estimation since they are all captured by the matching group fixed effects.

### F.4 Alternative Approaches to Dynamic Selection

In this section, we consider alternative approaches to address the dynamic selection problem. As discussed in the main text of the paper, the issue of dynamic selection arises when a user leaves the session right after the assignment to the treatment or control at period $t-1$. The key issue is that the data for these users are missing at the outcome collection phase $t$. Our main approach in the paper is to impute the specific ad that would have been shown at this missing exposure and assign the outcome zero to it. We now consider alternative approaches to impute the outcome.

| | *Dependent variable: Click ($Y^*_{i,t}$)* | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| Treatment ($W_{i,t}$) | 0.00221*** | 0.00225*** | 0.00198*** |
| | (9.77) | (9.69) | (9.59) |
| IPW Adjustment | ✓ | ✓ | ✓ |
| Imputed Sample | ✓ | ✓ | ✓ |
| Exposure ($t$) FE | ✓ | ✓ | ✓ |
| $Freq_{i,t}$ Indicators | ✓ | ✓ | ✓ |
| $Space_{i,t}$ Indicators | ✓ | ✓ | ✓ |
| $A^*_{i,t}$ FE | ✓ | ✓ | ✓ |
| $A_{i,t-1}$ FE | ✓ | ✓ | ✓ |
| $A^*_{i,t} \times L_{i,t-1}$ FE | | | ✓ |
| No. of Obs. | 2,405,695 | 2,405,695 | 2,405,695 |
| $R^2$ | 0.019 | 0.005 | 0.010 |
| Adjusted $R^2$ | 0.019 | 0.005 | 0.009 |
| *Note:* | | *p<0.05; **p<0.01; ***p<0.001 | |

Table W6: Average effects of the variety treatment on the CTR using different approaches for dynamic selection and imputation. Numbers reported in parentheses are t-statistics computed based on robust standard errors.

First, we consider a case where we predict what the outcome would have been had the user stayed in the session. This is similar to the practice in biostatistics literature, where the researcher uses the observed data to impute the missing observations (Little and Yau, 1996). As indicated in Challenge 2 in the main text of the paper, we need to impute $\{A^*_{i,t}, Y^*_{i,t}\}$. To do so, we first impute the ad $A^*_{i,t}$ using our approach. We then use a predictive approach to predict the click outcome on this ad had the user stayed in the session. For this purpose, we use an XGBoost classification algorithm, where we give it a large set of inputs and train it on the observed sample. Please see the approach in Rafieian and Yoganarasimhan (2021) for the details of features used in this prediction task. We then take the predictive model to the missing sample and obtain $\hat{Y}^*_{i,t}$ using our predictive algorithm. We impute the outcome ($Y^*_{i,t}$) as the predicted outcome $\hat{Y}^*_{i,t}$ from the XGBoost model. We then estimate the model in the main specification of the paper. The results are presented in the first column of Table W6. The estimated coefficients show the same pattern as before.

Second, we consider another outcome imputation technique based on the users' click decision

on the prior ad. A platform may want to distinguish between two types of leave events after the treatment assignment – (1) leave that is caused by the click on the ad shown at period $t-1$, and (2) leave that is not caused by click on the ad shown at period $t-1$. The former is still desirable for the platform, whereas the latter is the event the platform wants to avoid. To reflect this intuition in our outcome imputation, we impute the outcome as one if the user has left the session because of clicking on the ad. In other words, for the user who left at exposure $t-1$, if $Y_{i,t-1}=1$, then we have $Y^*_{i,t}=1$. We give this imputed outcome as the outcome of the main regression and estimate the treatment coefficients. The results shown in the second column of Table W6 reveal the same patterns – an increase in ad variety leads to an increase in CTR on the next ad.

Finally, we consider a case where we distinguish between ad fixed effects in the imputed vs. actual outcomes. This is because the user in the imputed case has not actually seen the ad. We take this account by controlling for the interaction of the current ad ($A^*_{i,t}$) and the leave decision at the prior exposure denoted by $L_{i,t-1}$. We re-estimate our model and present the results in the third column of Table W6. The estimated coefficient shows the same pattern as the main model.

### F.5 Clustering Adjustment in Standard Errors

We now discuss the issue of clustering in our estimation of standard errors in the main model. In light of Abadie et al. (2017), we know that a model should adjust for clustering in standard errors if: (1) assignment is clustered, and/or (2) sampling is clustered. Since our sample is the population of interest from a platform perspective, we mainly focus on the former type of clustering. Clustering in assignment means that for two clusters $C_1$ and $C_2$ in the data, there is a variance in their treatment assignment propensities. The most notable case is when the randomization is performed at the cluster level. That is, we first decide which clusters receive the treatment (either randomly or through some probabilistic distribution), and then all the users within the same cluster either receive the treatment or control, depending on the cluster assignment. As such, the issue of assignment clustering becomes less relevant if the randomization is performed at the observation level (Jones et al., 2019). This is the case in our empirical setting. Treatment assignment is performed at the impression level, as shown in Figure W3. In fact, the probabilistic allocation rule in our quasi-proportional auction prevents any level of clustering in the treatment assignment because all participating ads have a non-zero propensity of being shown.

However, this does not fully rule out the possibility of clustered treatment assignment. More formally, there is clustering in treatment assignment if the variance of treatment propensity for different clusters is non-zero. Hence, it only suffices to have $\Pr(W = 1 \mid C_j) \neq \Pr(W = 1 \mid C_k)$ for two clusters $C_j$ and $C_k$ in the data. Theoretically, this level of clustering may exist in our study at the level of all the inputs used to estimate the propensity scores in Web Appendix §D.1. However, in our IPW-adjusted regression, we obtain robust standard errors that incorporate all the variation in propensity scores, thereby reflecting clustering in treatment assignment (if any). Thus, we do not need to adjust for clustering in our main model.

Nevertheless, as a robustness check, we consider different plausible scenarios for clustered assignment and obtain clustered standard errors at the level corresponding to each scenario. We do that in addition to the regular robust standard errors in the IPW-adjusted regression model. In light of the discussion above, we expect to see no difference after accounting for different possible clustering in treatment assignment, since all the clustering is already taken into account. To find different plausible clustering in treatment assignment, we use Proposition 2 in the paper, which says that if two impressions share the same targeting characteristics and happen around the same time, their auctions are identical. For the treatment assignment, we know that we need to also take the prior set of ads into account. Thus, we consider the following clustering scenarios:

- Clustering at the interaction of $\textit{Target}_{i,t}$ and past variety $V_{i,t-1}$ (first column in Table W7).
- Clustering at the interaction of $\textit{Target}_{i,t}$ (second column in Table W7).
- Clustering at the interaction of Province and Hour-Day, since we know most of the discrepancy and clustering in assignment comes from this (third column in Table W7).
- Clustering at the propensity score level where levels are distinguished by 0.05 margin (fourth column in Table W7).

Together, all the results in Table W7 consistently show that at different levels of clustering, the estimated standard errors do not change. This is what we expected since robust standard errors in our IPW-adjusted regression takes the discrepancy and clustering in treatment assignment into account.

| | *Dependent variable: Click ($Y^*_{i,t}$)* | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Treatment ($W_{i,t}$) | 0.00186*** | 0.00186*** | 0.00186*** | 0.00186*** |
| | (8.38) | (8.45) | (8.48) | (8.87) |
| Clustering | *Target*$_{i,t} \times V_{i,t-1}$ | *Target*$_{i,t}$ | *Province* × *Hour* × *Day* | Propensity Score Level (0.05) |
| IPW Adjustment | ✓ | ✓ | ✓ | ✓ |
| Imputed Sample | ✓ | ✓ | ✓ | ✓ |
| Exposure ($t$) FE | ✓ | ✓ | ✓ | ✓ |
| *Freq*$_{i,t}$ Indicators | ✓ | ✓ | ✓ | ✓ |
| *Space*$_{i,t}$ Indicators | ✓ | ✓ | ✓ | ✓ |
| $A^*_{i,t}$ FE | ✓ | ✓ | ✓ | ✓ |
| $A_{i,t-1}$ FE | ✓ | ✓ | ✓ | ✓ |
| No. of Obs. | 2,405,695 | 2,405,695 | 2,405,695 | 2,405,695 |
| $R^2$ | 0.006 | 0.006 | 0.006 | 0.006 |
| Adjusted $R^2$ | 0.005 | 0.005 | 0.005 | 0.005 |
| *Note:* | | | | *p<0.05; **p<0.01; ***p<0.001 |

Table W7: Average effects of the variety treatment on the CTR with cluster-robust standards errors. Numbers reported in parentheses are t-statistics computed based on cluster-robust standard errors.

### F.6 Placebo Analysis

We now run a placebo analysis with a treatment that most likely has zero impact on the outcome. Our goal is to see if our model returns null effects for a treatment that likely has null effects. To find such a treatment, we use the same treatment variable of an increase in ad variety, but from the previous sessions. That is, if $i'$ is the last session where the user in session $i$ has been at exposure $t$, we define our placebo treatment to be $\textit{Placebo}_{i,t} = W_{i',t}$. The variable $\textit{Placebo}_{i,t}$ is missing if $W_{i',t}$ does not exist. We replace $W_{i,t}$ in our main specification with $\textit{Placebo}_{i,t}$ and run our model. In addition, we consider other specifications used in robustness check to ensure that those models are also robust to a placebo treatment. We present the results in Table W8. As expected, all the coefficients are null with very small t-statistics. Thus, our placebo analysis provides evidence that our main specification indeed delivers a null estimate for a potentially null effect.

| | *Dependent variable: Click ($Y^*_{i,t}$)* | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Placebo ($W_{i',t}$) | 0.00004 | −0.00000 | −0.00003 | 0.00013 |
| | (0.29) | (−0.03) | (−0.18) | (0.37) |
| IPW Adjustment | ✓ | ✓ | ✓ | ✓ |
| $A_{i,t-1}$ FE | ✓ | ✓ | | ✓ |
| $\textit{Freq}_{i,t}$ Indicators | ✓ | ✓ | | |
| $\textit{Space}_{i,t}$ Indicators | ✓ | ✓ | | |
| $A_{i,t-1}$ FE | ✓ | ✓ | | |
| $\textit{Target}_{i,t}$ FE | | ✓ | | |
| $V_{i,t-1}$ Indicators | | ✓ | | |
| $\textit{Post}_{i,t}$ FE | | | ✓ | |
| Matching Group FE | | | | ✓ |
| No. of Obs. | 2,000,639 | 2,000,639 | 2,000,639 | 2,000,639 |
| $R^2$ | 0.002 | 0.054 | 0.054 | 0.597 |
| Adjusted $R^2$ | 0.002 | 0.013 | −0.007 | −0.024 |
| *Note:* | | | | $^*$p<0.05; $^{**}$p<0.01; $^{***}$p<0.001 |

Table W8: Average effects of the placebo treatment on the CTR using different approaches. Numbers reported in parentheses are t-statistics computed based on robust standard errors.

# G Web Appendix G: Supplementary Results on the Mechanism

## G.1 Supplementary Materials for the Section on Treatment Effects Across Different Control Groups

In this section we first present the regression models used to generate Figure **??** in the paper and then show the robustness of those results to alternative specifications.

### G.1.1 Regression Tables for Figure 8

We run twelve separate regressions to generate Figure 8 – six separate regressions for Figure 8a and six separate regressions for Figure 8b. In all cases, we use the same treatment group $W_{i,t} = 1$, but different partitions of the control group ($W_{i,t} = 0$) based on their frequency and spacing. In Tables W9 and W10, we show the exact control group that we use to estimate the treatment effects for each column along with the estimates.

| | *Dependent variable: Click ($Y^*_{i,t}$)* | | | | | |
|---|---|---|---|---|---|---|
| | Control Condition: $Freq_{i,t-1} = k$ | | | | | |
| | (1) $k=1$ | (2) $k=2$ | (3) $k=3$ | (4) $k=4$ | (5) $k=5$ | (6) $k=6$ |
| Treatment ($W_{i,t}$) | 0.00128*** | 0.00247*** | 0.00383*** | 0.00628*** | 0.00699*** | 0.00604*** |
| | (5.58) | (7.02) | (7.34) | (8.97) | (6.66) | (3.49) |
| IPW Adjustment | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Imputed Sample | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Exposure ($t$) FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| $Freq_{i,t}$ Indicators | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| $Space_{i,t}$ Indicators | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| $A^*_{i,t}$ FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| $A_{i,t-1}$ FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| No. of Obs. | 1,879,546 | 1,544,301 | 1,379,397 | 1,310,135 | 1,276,831 | 1,260,508 |
| $R^2$ | 0.006 | 0.006 | 0.007 | 0.007 | 0.007 | 0.007 |
| Adjusted $R^2$ | 0.006 | 0.006 | 0.006 | 0.006 | 0.006 | 0.007 |
| *Note:* | | | | | | *p<0.05; **p<0.01; ***p<0.001 |

Table W9: Average effects of the variety treatment on the CTR when compared to the control group at different levels of past frequency ($Freq_{i,t-1}$). Numbers reported in parentheses are t-statistics computed based on robust standard errors.

| | *Dependent variable: Click* ($Y^*_{i,t}$) | | | | | |
|---|---|---|---|---|---|---|
| | Control Condition: $Space_{i,t-1} = l$ | | | | | |
| | (1) $l = 1$ | (2) $l = 2$ | (3) $l = 3$ | (4) $l = 4$ | (5) $l = 5$ | (6) $l = 6$ |
| Treatment ($W_{i,t}$) | 0.00278*** | 0.00184*** | 0.00212*** | 0.00107* | 0.00126* | 0.00032 |
| | (9.60) | (5.93) | (5.46) | (2.15) | (2.04) | (0.36) |
| IPW Adjustment | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Imputed Sample | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Exposure ($t$) FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| $Freq_{i,t}$ Indicators | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| $Space_{i,t}$ Indicators | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| $A^*_{i,t}$ FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| $A_{i,t-1}$ FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| No. of Obs. | 1,743,404 | 1,562,283 | 1,410,639 | 1,339,617 | 1,300,769 | 1,278,123 |
| $R^2$ | 0.006 | 0.007 | 0.007 | 0.007 | 0.007 | 0.007 |
| Adjusted $R^2$ | 0.006 | 0.006 | 0.006 | 0.006 | 0.006 | 0.006 |
| *Note:* | | | | | | *p<0.05; **p<0.01; ***p<0.001 |

Table W10: Average effects of the variety treatment on the CTR when compared to the control group at different levels of spacing ($Space_{i,t-1}$). Numbers reported in parentheses are t-statistics computed based on robust standard errors.

### G.1.2 Robustness Checks to test Predictions 1 and 2

We now focus on an alternative modeling approach to check the robustness of the results shown in Figure 8. Instead of running separate regressions comparing the treatment with different control groups based on $\textit{Freq}_{i,t-1}$ and $\textit{Space}_{i,t-1}$, we can directly include these variables in the model. However, we must notice that our propensity scores are estimated for the binary treatment $W_{i,t}$. As such, we need to either control for the propensity scores of these new variables or control for the variables determines propensity scores of $\textit{Freq}_{i,t-1}$ and $\textit{Space}_{i,t-1}$. We take the latter approach and control for $\textit{Target}_{i,t}$, since we have shown that two sessions with the same targeting characteristics that happen around the same time have the same ad allocation distribution (Proposition 2). If the ad allocation distribution (i.e., auction) is identical for two sessions, then the assignment to $\textit{Freq}_{i,t-1}$ and $\textit{Space}_{i,t-1}$ is as good as random.

We estimate two models using $\textit{Freq}_{i,t-1}$ and $\textit{Space}_{i,t-1}$ as covariates and present the results in Table W11. Overall, the results shown in this table reveal consistent patterns with our results in Figure 8. The estimated coefficient for $\textit{Freq}_{i,t-1}$ is negative, which is in line with Prediction 1: the more frequently the ad has been shown in the past, the less likely it is for the user to click on the ad shown after that. On the other hand, as predicted by Prediction 2, the coefficient for $\textit{Space}_{i,t-1}$ is positive. The greater the spacing is at point $t-1$, the higher the likelihood of clicking on the ad shown at $t$.

| | *Dependent variable: Click ($Y^*_{i,t}$)* | |
|---|---|---|
| | (1) | (2) |
| $\textit{Freq}_{i,t-1}$ | $-0.00072^{***}$ | |
| | $(-7.84)$ | |
| $\textit{Space}_{i,t-1}$ | | $0.00025^{***}$ |
| | | (5.52) |
| Exposure ($t$) FE | ✓ | ✓ |
| $\textit{Freq}_{i,t}$ Indicators | ✓ | ✓ |
| $\textit{Space}_{i,t}$ Indicators | ✓ | ✓ |
| $A^*_{i,t}$ FE | ✓ | ✓ |
| $A_{i,t-1}$ FE | ✓ | ✓ |
| $\textit{Target}_{i,t}$ FE | ✓ | ✓ |
| No. of Obs. | 2,405,695 | 2,405,695 |
| $R^2$ | 0.056 | 0.056 |
| Adjusted $R^2$ | 0.019 | 0.019 |
| *Note:* | $^{*}$p<0.05; $^{**}$p<0.01; $^{***}$p<0.001 | |

Table W11: Average effects of $\textit{Freq}_{i,t-1}$ and $\textit{Space}_{i,t-1}$ on the CTR on the next ad. Numbers reported in parentheses are t-statistics computed based on robust standard errors.

## G.2 Supplementary Materials for the Section on Heterogeneity Across Past Usage Frequency and Recency

In this section we first present the regression models used to generate Figure 9 in the paper and then show the robustness of those results to alternative specifications, where we use interactions.

We present the results of Figure 9 in Table W12. These are the models used in the main text to generate the results shown in Figure 9. As such, for each column, we focus on a specific sub-sample of the data.

| | *Dependent variable: Click ($Y^*_{i,t}$)* | | | |
|---|---|---|---|---|
| | Data Split | | | |
| | (1) Low Usage Frequency | (2) High Usage Frequency | (3) Low Usage Recency | (4) High Usage Recency |
| Treatment ($W_{i,t}$) | 0.00250*** | 0.00022 | 0.00301*** | 0.00041 |
| | (7.51) | (1.03) | (9.29) | (1.64) |
| IPW Adjustment | ✓ | ✓ | ✓ | ✓ |
| Imputed Sample | ✓ | ✓ | ✓ | ✓ |
| Exposure ($t$) FE | ✓ | ✓ | ✓ | ✓ |
| $Freq_{i,t}$ Indicators | ✓ | ✓ | ✓ | ✓ |
| $Space_{i,t}$ Indicators | ✓ | ✓ | ✓ | ✓ |
| $A^*_{i,t}$ FE | ✓ | ✓ | ✓ | ✓ |
| $A_{i,t-1}$ FE | ✓ | ✓ | ✓ | ✓ |
| No. of Obs. | 1,336,631 | 1,069,064 | 1,246,626 | 1,159,069 |
| $R^2$ | 0.009 | 0.002 | 0.009 | 0.003 |
| Adjusted $R^2$ | 0.008 | 0.001 | 0.008 | 0.003 |
| *Note:* | | | | *p<0.05; **p<0.01; ***p<0.001 |

Table W12: Heterogeneity in the variety effects across usage frequency and recency. Numbers reported in parentheses are t-statistics computed based on robust standard errors.

Next, we consider a regression model where we use the entire sample and identify the heterogeneity in the main effects by estimating the interaction coefficients. To do that, we formally define both variables $UsageFreq_i$ and $UsageGap_i$. As defined in the main text of the paper, $UsageFreq_i$ is the number of impressions the user has seen in the past prior to the current session and $UsageGap_i$ is the time between the last impression in the previous session and the first impression in the current session in terms of hours. Given the skewed distribution of both variables $UsageFreq_i$ and $UsageGap_i$, we take their logs and include them in the regression model with interactions.

We present the results for these alternative specifications in Table W13. The estimated coefficients for interactions reveal the same insights as Figure 9. Most notably, in the fourth column of this model, we notice the same patterns when we control for user fixed effects. Controlling for user fixed effects helps us focus exclusively on within-user variation. That is, our model only uses the variation within each user to estimate the main effects. As shown in the estimates presented in column (4), the results remain the same, suggesting that the source for this heterogeneity is the time-varying user characteristics (e.g., the same user when she has seen fewer vs. more impressions).

| | *Dependent variable: Click ($Y^*_{i,t}$)* | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Treatment ($W_{i,t}$) | 0.00331*** | 0.00071** | 0.00220** | 0.00171* |
| | (4.70) | (2.71) | (3.09) | (2.35) |
| $\log(\textit{UsageFreq}_i)$ | −0.00703*** | | −0.00488*** | −0.00740*** |
| | (−67.27) | | (−51.33) | (−55.26) |
| $\log(\textit{UsageGap}_i)$ | | 0.00766*** | 0.00409*** | 0.00431*** |
| | | (54.56) | (29.05) | (29.73) |
| Treatment × $\log(\textit{UsageFreq}_i)$ | −0.00050*** | | −0.00035** | −0.00045** |
| | (−3.51) | | (−2.75) | (−3.39) |
| Treatment × $\log(\textit{UsageGap}_i)$ | | 0.00059** | 0.00037* | 0.00043* |
| | | (3.04) | (1.99) | (2.22) |
| IPW Adjustment | ✓ | ✓ | ✓ | ✓ |
| Imputed Sample | ✓ | ✓ | ✓ | ✓ |
| Exposure ($t$) FE | ✓ | ✓ | ✓ | ✓ |
| $\textit{Freq}_{i,t}$ Indicators | ✓ | ✓ | ✓ | ✓ |
| $\textit{Space}_{i,t}$ Indicators | ✓ | ✓ | ✓ | ✓ |
| $A^*_{i,t}$ FE | ✓ | ✓ | ✓ | ✓ |
| $A_{i,t-1}$ FE | ✓ | ✓ | ✓ | ✓ |
| User FE | | | | ✓ |
| No. of Obs. | 2,405,695 | 2,405,695 | 2,405,695 | 2,405,695 |
| $R^2$ | 0.019 | 0.005 | 0.010 | 0.086 |
| Adjusted $R^2$ | 0.019 | 0.005 | 0.009 | 0.058 |
| *Note:* | | | | *p<0.05; **p<0.01; ***p<0.001 |

Table W13: Heterogeneity in the variety effects across usage frequency and recency using interactions. Numbers reported in parentheses are t-statistics based on robust standard errors.

# References


Abadie, A., Athey, S., Imbens, G. W., and Wooldridge, J. (2017). When should you adjust standard errors for clustering? Technical report, National Bureau of Economic Research.

Austin, P. C. (2009). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in medicine*, 28(25):3083–3107.

Godes, D. and Mayzlin, D. (2004). Using Online Conversations to Study Word-Of-Mouth Communication. *Marketing Science*, 23(4):545–560.

Jones, D., Molitor, D., and Reif, J. (2019). What do workplace wellness programs do? evidence from the illinois workplace wellness study. *The Quarterly Journal of Economics*, 134(4):1747–1791.

Little, R. and Yau, L. (1996). Intent-to-treat Analysis for Longitudinal Studies with Drop-outs. *Biometrics*, pages 1324–1333.

McCaffrey, D. F., Griffin, B. A., Almirall, D., Slaughter, M. E., Ramchand, R., and Burgette, L. F. (2013). A tutorial on propensity score estimation for multiple treatments using generalized boosted models. *Statistics in medicine*, 32(19):3388–3414.

Rafieian, O. and Yoganarasimhan, H. (2021). Targeting and Privacy in Mobile Advertising. *Marketing Science*, 40(2):193–218.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT press.