MARKETING SCIENCE

Effectiveness of Advertising Campaigns on Short-Form Video Social Platforms: An Empirical Analysis through a Large-Scale Randomized Field Experiment on ByteDance

Journal:	*Marketing Science*
Manuscript ID	MKSC-2023-0575.R1
Manuscript Type:	Original Article
Keywords:	Advertising, Field experiments, Platform

Abstract:	Short-form videos have taken over social media and attracted attention from advertisers. As brands shift their advertising spending to short-form video social platforms, doubts remain about the advertising efficacy on these platforms. In a large-scale randomized experiment on ByteDance in collaboration with an automobile brand, we show a significant effect of an advertising campaign on ByteDance. Most of the advertising effect comes from advertising spillover beyond ByteDance, with exposed users being eight times more likely to convert from outside than from within ByteDance, which raises the importance of information sharing between the platform and brands. When considering conversions outside ByteDance, the average cost per conversion, which brands commonly use to evaluate the cost of campaigns, shrinks by 5 or 25 times, depending on the methods used to calculate it. Information sharing can also affect a brand's targeting strategy. While commonly used demographic variables by the automobile brand are effective for target marketing with only platform data, they are not when considering conversions outside ByteDance. Instead, a behavioral variable proposed herein (prior brand home page visits) effectively moderates the advertising effect but has no impact when including only platform data in the analysis.

Effectiveness of Advertising Campaigns on Short-Form Video Social Platforms: An Empirical Analysis through a Large-Scale Randomized Field Experiment on ByteDance

Abstract

Short-form videos have taken over social media and attracted attention from advertisers. As brands shift their advertising spending to short-form video social platforms, doubts remain about the advertising efficacy on these platforms. In a large-scale randomized experiment on ByteDance in collaboration with an automobile brand, we show a significant effect of an advertising campaign on ByteDance. Most of the advertising effect comes from advertising spillover beyond ByteDance, with exposed users being eight times more likely to convert from outside than from within ByteDance, which raises the importance of information sharing between the platform and brands. When considering conversions outside ByteDance, the average cost per conversion, which brands commonly use to evaluate the cost of campaigns, shrinks by 5 or 25 times, depending on the methods used to calculate it. Information sharing can also affect a brand's targeting strategy. While commonly used demographic variables by the automobile brand are effective for target marketing with only platform data, they are not when considering conversions outside ByteDance. Instead, a behavioral variable proposed herein (prior brand home page visits) effectively moderates the advertising effect but has no impact when including only platform data in the analysis.

Keywords: digital advertising, ad platform, ad spillover, advertising effect

1. Introduction

Short-form videos have taken over social media in recent years, popularized by TikTok and quickly copied by Instagram Reels and YouTube Shorts. These videos, which are bite-sized, digestible, highly addictive, and often under 60 seconds long, easily catch the attention of time-strapped audiences. They are also readily sharable and easy to create, which makes them popular in social networks. As of 2023, short-form video platforms such as YouTube Shorts had 1.5 billion monthly active users, and TikTok and Instagram Reels had more than 2 billion combined (Shore 2023).

Brands have quickly recognized the value of short-form videos. According to HubSpot (2023), 73% of consumers prefer to watch short-form videos to learn about products or services. In addition, 58% of viewers indicate that they will watch an entire business video if it is less than 60 seconds long, because they view short-form videos as 2.5 times more engaging than long-form ones (Vidyard 2023). As a result, brands have shifted much of their advertising budget to short-form video platforms. In a survey among more than 1000 global marketers, 33% plan to invest most in short-form video platforms such as TikTok and Instagram Reels (HubSpot 2023).

However, some brands are skeptical about the effectiveness of short-form video platforms. For example, they experience lower conversion rates on these platforms than in other outlets, such as in our case. Thus, advertising' effectiveness on digital ad platforms continues to be a concern (Aral 2021). In past years, firms such as General Motors and Procter & Gamble have pulled back ad spending on big digital platforms, citing minimal impact on consumers' purchases (Terlep et al. 2012, Vranica 2018). Research studies also reveal relatively low advertising effects through randomized experiments. For example, Blake et al. (2015) found that Google search engine marketing had a small and nonsignificant effect on sales in a large-scale experiment. Johnson et al. (2017b) reported the median lift in conversion as 17% for site visits and 8% for conversions in 432 experiments on the Google Display Network. Gordon et al. (2023) found that the median lift in conversion was only 9% among 663 randomized controlled trials with a mix of display and video ads[1] run on Facebook. Will the same problem happen on short-form video platforms? This question is particularly important for brands planning to increase spending on these social platforms. At the same time, short-form video platforms also want to know the answer to justify their skyrocketing advertising prices in recent years. For example, reports estimate that the cost per thousand (CPM) on TikTok increased by 92% from 2020 to 2021 (Rosenfeld 2022).

However, research on the advertising effect of short-form videos is scant; an exception is Yang et al. (2021), who assess influencer video advertising on TikTok. Although influencer video advertising is a fast-growing market, sponsor ads directly made by advertisers constitute the majority of short-form video advertising.[2] In this study, we aim to examine the advertising effectiveness on a short-form video platform. Considering the difficulty in measuring digital advertising effects (Gordon et al. 2021), we follow recent studies (Gordon et al. 2019, 2023) and conduct a large-scale randomized field experiment on ByteDance in collaboration with an automobile brand. The brand has invested heavily in ByteDance and plans to shift all its digital advertising budget to the platform; at the same time, however, it is skeptical of the effectiveness of advertising on ByteDance. More than 84 million users were randomly assigned to control and treatment groups, and only those in the treatment groups were exposed to the short-form video ads. Conversion refers to whether users provided contact information such as names and cell phone numbers to the brand. The experiment lasted for seven weeks and took another two weeks for completion of conversions. To address its question (i.e., the effect if the brand advertises only on ByteDance), the automobile brand ceased all digital advertising campaigns outside ByteDance during the experiment period and carried no digital advertising campaign in the two weeks after the experiment. We matched the ByteDance users in the experiment and the brand's converted customers using their cellphone numbers. This enabled us to identify the conversions within and outside ByteDance for users in the experiments.

[1] This information was not included in the original paper. It was kindly provided by one of the authors upon our request.
[2] In 2022, the US influencer marketing spending on TikTok was $774.8 million (Gutelle 2022), and TikTok's US ad revenue was $9.9 billion (Winter 2023).

The results show a significant effect of the short-form video advertising campaign on ByteDance. The advertising effect measured by ATT (average treatment effect on the treated) is 0.082‰, and the ATT lift on conversion is 1.1226, which is much higher than the advertising effects reported in various experiments with mostly display ads (Johnson et al. 2017b, Gordon et al. 2023). Of note, most of the advertising effect comes from advertising spillover[3] beyond ByteDance. In the experiment, the exposed users were more than eight times likely to convert from outside than from within ByteDance. Considering the conversion outside ByteDance, the advertising effect measured by ATT is four times that calculated using only platform data (conversions within ByteDance). As social platforms and brands typically do not share their customer information, this finding calls for caution by brands when calculating the advertising costs. In our context, brands often use cost per conversion (CPC) for the economic evaluation of advertising campaigns. Using brands' commonly adopted method to calculate CPC, we show that it shrinks by 25 times when considering conversion outside ByteDance. When calculating CPC with the ATT, a conceptually correct method, CPC reduces to 20% when including the outside conversion.

Surprisingly, we find that sharing information between the platform and brand also has a profound implication for the target marketing strategy. In China, automobile brands often rely on several key demographic variables for targeting to reach a wide range of customers. Our results show that these variables are effective when using only platform data for analysis. However, when including conversions outside ByteDance, the variables are no longer effective. This finding implies a misinformed targeting strategy if the platform and brand do not share customer information. To assist the brand, we propose the use of behavioral variables for targeting and demonstrate that with users' prior brand home page visits on Douyin (the Chinese version of TikTok). We show that this variable moderates the advertising effect when complete conversion information is used and thus can be employed for targeting. However, when we include only platform data for analysis, this variable becomes ineffective for targeting.

[3] We broadly define advertising spillover as advertising in one channel affecting consumer behavior in other channels.

These findings are managerially important. They not only mitigate brands' concerns about the advertising effect on short-form video platforms but also help these platforms better communicate with their clients. However, the significant advertising spillover discovered herein underscores the importance of information sharing between the platform and brands. Without that, neither the platform nor brands can correctly evaluate the advertising effect, which is the base to price advertising, or design a proper targeting strategy. The commonly used "conversion pixel" (Johnson et al. 2017b, Gordon et al. 2023), which is a piece of code ad platforms provide for brands to insert in specific web pages to record users' page views, is of limited use when conversions occur in channels out of a brand's control, because the brand cannot force those channels to insert the conversion pixel on their own web pages.

2. Institutional Details

2.1. Advertising on ByteDance

ByteDance is a global leading internet company headquartered in China. It has two prominent applications in China, Douyin and Toutiao. Douyin is the Chinese counterpart of TikTok, and Toutiao is a news aggregator similar to Google News. In a typical advertising campaign on ByteDance, ads often appear on both Douyin and Toutiao.

On ByteDance, ads can be either display (i.e., images) or short-form videos. In recent practice, the majority of ads on ByteDance have been short-form videos. Thus, the ads used in our experiment are only short-form video ads. On ByteDance, ads appear either in newsfeeds or on open screens. Newsfeed ads are included in streams of short videos or text newsfeeds, while open-screen ads appear upon the launch of the app. Newsfeed ads are typically 10 to 30 seconds long, while open-screen ads are typically 3 or 5 seconds long. The price of ads can be either fixed or determined by auction. A fixed price is an agreement-based contract in which the advertiser and the platform negotiate on several critical elements before the campaign, such as the campaign period, total exposure, and price. Auction entails a real-time bidding process, which is commonly used on most internet platforms. During our experiment, ByteDance used a fixed price for open-screen ads and either a fixed price or auction for newsfeed ads.

Figure 1a is an example of an open-screen video ad on Douyin for an automobile advertiser.[4] The left panel is a screen shot of the short-form video played when users launch the Douyin app. Users can skip the ad by clicking on the upper-right-hand button or simply let the video finish, after which they will enter the app's main page. If users swipe up the ad for more details, they will be directed to the middle panel, from which an additional swipe-up will lead them to the landing page, shown in the right panel. On the landing page, users can book an appointment with a dealer by filling out the form with their names and cell phone numbers for sales agents to contact them afterward.[5]

Douyin serves users with a stream of short-form videos according to its recommendation algorithm, and some slots are reserved for advertisements. Figure 1b illustrates an example of a newsfeed video ad on Douyin for an automobile advertiser. The left panel is a screen shot of the short-form video commercial. Users can either swipe up to skip the ad and watch the next video or click on the associated button for more details, which will then direct them to the middle panel. With another swipe up, they will go to the landing page, as shown in the right panel, on which they can book an appointment by filling out the form. Similarly, Figure 1c illustrates video ads of an automobile brand appearing in the newsfeed on Toutiao.

2.2. Ad Delivery Process on ByteDance

Advertisers first select their targeting criteria so that ads will only be sent to relevant users. In the automobile industry in China, rather than using targeting strategies with many fine-grained variables, advertisers typically focus on a few demographic variables to ensure a broad reach to users on the platform. Commonly used targeting variables include gender, tier of city, and age. Ads are then delivered to users satisfying the criteria set by the advertiser.

The ad delivery process for open-screen ads differs from that for newsfeed ads. The actual process is complicated, involving technical details proprietary to the company, such as optimization of

[4] Owing to confidentiality agreements, all figural examples shown herein do not come from the brand studied. They are for illustration purposes only.
[5] Such practice is common in the China market. However, owing to the privacy concern, most users only provide their last names rather than their full names.

data transfer, mobile cache usage, and waiting time. Thus, what we describe herein is only somewhat conceptually equivalent to the actual engineering process.

The delivery of open-screen ads follows four stages: preload, trigger, selection, and display. In the preload stage, a batch of ads are selected and preloaded to users' cache to minimize the waiting time. The selection of ads is mainly determined by user and ad features, as well as contract specifics, such as price. The number of ads to be selected is not limited, but empirically, it is often less than 10 ads in total. When a user launches the app, an advertising slot—the open screen—is triggered. Then, the platform selects an ad from the batch of preloaded ads. In most cases, the selection is random with equal probabilities. Finally, in the display stage, the selected ad from the previous stage is retrieved from the cache and displayed to the user.

By contrast, the sequence of actions in the delivery process of newsfeed ads is trigger, selection, preload, and display. Whenever a user swipes for more content, it triggers an advertising slot. The platform will load certain types of content (e.g., videos on Douyin, news on Toutiao); one slot is typically reserved for an advertisement, and the others are called "natural content." In the selection stage, a winning ad will be chosen from the pool of fixed-price ads randomly with equal probability. At the same time, another winning ad will be determined by auction for ads offering bids. The final winning ad will be decided between these two ads primarily depending on their prices. Afterward, the algorithm will rank the content (typically 10 slots including one ad) to optimize user experience and preload the content to the user's cache. Finally, the ad will be displayed when the user browses to its position. However, if the user leaves the app before reaching the ad slot, the ad will not be displayed, and the cache will be cleared.

3. Experimental Design

In line with Gordon et al. (2023), the ad exposure on ByteDance is endogenously determined by selections induced by user activity, targeting, and competition. Therefore, we conduct a large-scale randomized experiment to estimate the effects of a short-form video advertising campaign on ByteDance. As mentioned previously, the experiment is in collaboration with ByteDance and a leading automobile brand.[6] In the Chinese market, the automobile industry contributed more than 10% of advertising expenditure in 2022.[7] Given the limited number of firms, automobile brands are generally key accounts for advertising platforms. The brand has invested heavily on ByteDance and was considering shifting all its digital advertising budget to it. However, the brand was also concerned about the effectiveness of advertising on ByteDance because it observed a lower conversion rate on ByteDance than in other advertising channels, such as vertical automobile websites. This is especially concerning given the much higher advertising price on ByteDance than in the other channels.

The experiment was managed by "Ocean Engine," the digital marketing service division of ByteDance. In the experiment, the brand was promoting a new sedan on ByteDance. The campaign was run on both Douyin and Toutiao and used only short-form video ads.

In our context, a user refers to a real person. While some users may have multiple accounts, ByteDance is able to identify these users through proprietary methods. Notably, the brand's main objective in the experiment was to identify the advertising effect if it concentrated all its digital advertising budget on ByteDance. Thus, the brand ceased all digital advertising campaigns other than that used in the experiment,[8] according to the following schedule:

1. *Stage 1* (August 13, 2020–September 30, 2020): experiment on ByteDance, no digital advertising campaign outside ByteDance.
2. *Stage 2* (October 1, 2020–October 14, 2020): no digital advertising campaign within or outside ByteDance.

We randomly assigned the focal advertiser's targeted users to either the treatment or control group. After that, Ocean Engine executed the ad delivery process to deliver the focal ads to these targeted users. We use diagrams to describe the experiment procedure. As Figure 2a shows, in terms of the ad delivery process for users in the treatment group, no experimental intervention occurred (see Section 2.2).

[6] Confidentiality agreement prevents us from revealing the brand's identity.
[7] https://baijiahao.baidu.com/s?id=1754822274656302264&wfr=spider&for=pc (in Chinese).
[8] The brand's offline advertising campaigns remained unchanged during the experiment period.

For users in the control group, interventions occurred during the ad delivery process. As Figure 2b shows, for open-screen ads, when the focal ad was about to preload to the targeted users' cache, the system simply blocked it from the batch of the selected ads; for newsfeed ads, if the focal ad won in the selection stage, the system reran the selection stage without the focal ad to find a new winner (the next-best ad) to replace the focal ad.

Even in the treatment group, the focal ads may not have been preloaded to the cache for some users. This happened in newsfeed ads when the focal ads were not chosen in the selection stage. The same scenario could also happen in the control group, in which the subsequent blocking of the focal ads becomes unnecessary. As the system keeps the complete log of the ad delivery process, for all targeted users, we know whether they were actually preloaded with the focal ads in the treatment condition or supposed to be preloaded with the focal ads in the control condition. Such information is critical for our analyses, which we discuss subsequently.

In terms of exposure, users in the control group were not exposed to the focal ads by design. However, users in the treatment group may or may not have been exposed, owing to the compliance problem. For example, for open-screen ads, the eventual display of the focal ads depends on the usage frequency of the apps (i.e., frequent usage leads to a greater chance of exposure); for newsfeed ads, if users do not browse to the focal ad's position, the ad will not be displayed. As compliance is endogenous, it generates systematic difference between the exposed and unexposed users in the treatment group (Gordon et al. 2019).

Finally, the conversion in our experiment is the acquisition of customer leads, which entails users providing their contact information (e.g., names,[9] cell phone numbers) for the brand to reach out to them with more product information. As automobile purchase is an important decision for many people, they often search for product information online. Automobile brands seize on the opportunity to obtain contact information in exchange for more product information. After acquiring leads, salespeople then contact the

[9] For privacy concerns, users often leave only their first names or an alias.

potential customers with the purpose of having them visit the dealer store. After people are in the store, the eventual purchase is determined by the product quality, sales effort, and customer service, rather than the early advertising. In practice, the sales conversion rate of in-store customer leads is fairly high (approximately 30%), according to industry experts. Thus, collecting customer leads becomes a major marketing strategy for automobile brands in China. Automobile brands often treat acquiring customer leads as conversion in the advertising campaign and use this conversion as the key metric to evaluate the advertising campaign, rather than final sales.

Conversion could occur either on ByteDance, when users click on the enclosed link in the focal ads and fill out the form to leave their contact information, or in other digital channels, when users search for the car information later. Other digital channels outside ByteDance include automobile-specialized apps, the brand's own channels, and other digital media. In the experiment, the majority of conversions outside ByteDance came from the automobile-specialized apps; major ones are Autohome, DongCheDi, and Yiche in the China market. These apps provide various automobile-related content and extensive automobile-listing information to consumers and therefore are the first stop for consumers who are interested in cars or want to purchase one. Take Autohome as an example. When consumers search for car information and land on the page of a specific car model, as illustrated in Figure 3, they can click on the associated button to contact a sales representative. Because Autohome users need to register with their cell phone numbers to enable such a live-chat function, the brand acquires their cell phone numbers upon initiation of a live-chat session. In that case, a conversion occurs. Alternatively, a conversion event in Autohome can occur when users consult the floor price for this model by providing their contact information. The process is similar in other automobile-specialized apps. In addition to these apps, a small proportion of conversions occur in the brand's own channels or on other digital media. The brand's own channels consist of its own app and social media accounts, with its app accounts for most conversions in this experiment. The other digital media include Tencent, Sina, and NetEase, which often have sections for automobile content. For example, there is a section for car content on NetEase (www.163.com), in which customers can visit pages related to our focal brand. Similar to the automobile-specialized apps, on those pages, customers can also contact the brand's representative or consult the price, leading to conversions.

4. Analysis of the Experiment

We first develop the causal framework of the advertising effect for our experiment. The estimation method mainly follows that of Imbens and Angrist (1994) and Gordon et al. (2019).

4.1. Definitions and Assumptions

We denote $Z_i \in \{0, 1\}$ as the treatment status of user *i*, where 0 is the control and 1 is the treatment, and denote $PL_i(Z_i) \in \{0, 1\}$ as the preload status of user *i*, which depends on the treatment status. In our experimental procedure, we have $PL_i(1) \in \{0, 1\}$; that is, if user *i* is assigned to the treatment group, the focal ad may or may not preload to his or her cache; however, $PL_i(0) = 0$, because if user *i* is assigned to the control group, the focal ad will never be preloaded to his or her cache. We also denote $W_i(Z_i, PL_i) \in \{0, 1\}$ as an indicator of whether the focal ad was eventually displayed to user *i*. Moreover, $W_i(0, PL_i(0)) = W_i(0, 0) = 0$ for all *i*; that is, users in the control group are not exposed to the focal ad because the ad is not preloaded to their cache. In addition, when $PL_i(1) = 0$, we have $W_i(1, PL_i(1)) = W_i(1, 0) = 0$; that is, users in the treatment group, who are not preloaded with the focal ad, are not exposed to the ad. However, when $PL_i(1) = 1$, $W_i(1, PL_i(1)) = W_i(1, 1) \in \{0, 1\}$; that is, users in the treatment group, who are preloaded with the focal ad, are either exposed to the ad or not. Finally, we denote $Y_i(Z_i, PL_i, W_i) \in \{0, 1\}$ as the potential conversion outcome, where 1 is conversion and 0 otherwise. For example, $Y_i(1, 1, 1) = 1$ indicates that if user *i* is assigned to the treatment group, preloaded and exposed to the focal ad, he or she could convert during the conversion period.

We make three standard assumptions for valid inference. First, *stable unit treatment value assumption* means that a user can receive only one version of the treatment, and this does not interfere with another user's outcomes. In our experiment, ByteDance's ability to identify each user ensures a single treatment to each user in the experiment. As the experiment was unknown to the users, it is unlikely that they exchanged information on this subject. If the exposed users in the treatment group indeed shared ads to the users in the control group, the test of the advertising effect would be a conservative one (Gordon et al. 2019).

Second, *random assignment* means that the assignment of the treatment is random across users or that the distribution of Z_i is independent of all potential outcomes, preload status, and exposure conditions, that is, $\{Y_i\big(0, PL_i(0), W_i(0)\big), Y_i\big(1, PL_i(1), W_i(1)\big)\} \perp Z_i$, $\{PL_i(0), PL_i(1)\} \perp Z_i$, $\{W_i\big(0, PL_i(0)\big), W_i\big(1, PL_i(1)\big)\} \perp Z_i$. In other words, whether a user will be preloaded with the focal ad, be exposed to the focal ad, or convert in a condition (treatment/control) is irrelevant to which condition this user is assigned in the experiment. Note that although the assignment through Z_i is random, the preloading and exposure are not necessarily random due to the compliance problem. The randomization of the treatment is untestable because we do not observe all potential outcomes, preloading status, or exposure conditions. For example, for users in the treatment group, we cannot observe whether they would have converted had they been assigned to the control group. Similarly, we cannot observe whether users in the control group would have been preloaded with the focal ads, been exposed to it, or converted had they been assigned to the treatment group. Following the conventional approach, we performed randomization checks and found no evidence against proper randomization.

Third, *exclusion restriction* means that the assignment affects the outcome only through the ad exposure; that is, $Y_i(0, PL(0), w) = Y_i(1, PL(1), w), \forall\, w \in \{0,1\}$. In our context, the exclusion restriction indicates that what truly matters to the user's conversion is ad exposure, not the assignment to the treatment or control group. Our experimental procedure ensures that the assignment only affects ad exposure with no direct influence on users' conversion behavior.

4.2. Advertising Effect

Following the literature on digital advertising (Johnson et al. 2017a, Gordon et al. 2019), we define the advertising effect as the average treatment effect on the treated (ATT),

$$ATT = E\left[Y\left(1, PL(1),\, W\big(1,\, PL(1)\big)\right) - Y\left(0, PL(0),\, W\big(0,\, PL(0)\big)\right) \middle| W\big(1,\, PL(1)\big) = 1\right], \quad (1)$$

where W(1, PL(1)) = 1 means that the ATT is conditional on users who end up exposed to the focal ad had they been assigned to the treatment group. According to our previous discussion, Equation (1) is equivalent to

$$ATT = E\left[Y\left(1, PL(1),\, W\big(1,\, PL(1)\big)\right) - Y\left(0, PL(0),\, W\big(0,\, PL(0)\big)\right)\Big|\, W\big(1,\, PL(1)\big) = 1,\, PL(1) = 1\right].$$

(2)

This is because the necessary condition for W(1, PL(1)) = 1 is PL(1) = 1; that is, the focal ad must first be preloaded to the user's cache to be displayed. Following Imbens and Rubin (2015), we can reexpress the ATT in Equation (2) in the following way. We first define the intent to treat conditional on preload ($\mathrm{ITT_{PL}}$) as

$$ITT_{PL} \equiv E\left[Y\left(1, PL(1),\, W\big(1,\, PL(1)\big)\right) - Y\left(0, PL(0),\, W\big(0,\, PL(0)\big)\right)\Big|\, PL(1) = 1\right], \qquad (3)$$

and then we have

$$ITT_{PL}$$

$$= E\left[Y\left(1, PL(1),\, W\big(1,\, PL(1)\big)\right) - Y\left(0, PL(0),\, W\big(0,\, PL(0)\big)\right)\Big|\, W\big(1,\, PL(1)\big) = 1,\, PL(1) = 1\right]$$

$$\cdot \Pr\left[W\big(1,\, PL(1)\big) = 1 \,\big|\, PL(1) = 1\right]$$

$$+E\left[Y\left(1, PL(1),\, W\big(1,\, PL(1)\big)\right) - Y\left(0, PL(0),\, W\big(0,\, PL(0)\big)\right)\Big|\, W\big(1,\, PL(1)\big) = 0,\, PL(1) = 1\right]$$

$$\cdot \Pr\left[W\big(1,\, PL(1)\big) = 0 \,\big|\, PL(1) = 1\right]$$

$$= \rho \cdot ATT + (1-\rho) \cdot E\left[Y\left(1, PL(1),\, W\big(1,\, PL(1)\big)\right) - Y\left(0, PL(0),\, W\big(0,\, PL(0)\big)\right)\Big|\, W\big(1,\, PL(1)\big) = 0,\, PL(1) = 1\right]$$

$$= \rho \cdot ATT, \qquad (4)$$

where $\rho = \Pr\left[W\big(1,\, PL(1)\big) = 1 \,\big|\, PL(1) = 1\right]$ is the compliance ratio, or the probability that users end up being exposed to the focal ad had they been assigned to the treatment group and preloaded with the focal ad. The first equality of Equation (4) holds by definition. The third holds because of the exclusion restriction assumption; that is, Y(1, PL(1), 0) = Y(0, PL(0), 0). In summary, we have

$$\pi \equiv ATT = \frac{ITT_{PL}}{\rho}. \tag{5}$$

Finally, we define the ATT lift, π_l, as

$$\tau_l \equiv \frac{\pi}{\mathrm{E}\left[Y\left(0, PL(0), W\left(0, PL(0)\right)\right) \middle| W\left(1, PL(1)\right)=1, PL(1)=1\right]}$$

$$= \frac{\pi}{\mathrm{E}\left[Y\left(1, PL(1), W\left(1, PL(1)\right)\right) \middle| W\left(1, PL(1)\right)=1, PL(1)=1\right] - \pi}. \tag{6}$$

The ATT lift measures the advertising effect (i.e., π) as a proportion to the baseline, which is the conversion without ad exposure.

Regarding the advertising effect, several points are worth further explanation. First, the reason to focus on users to whom the brand's ad would be preloaded and exposed had the brand ran the campaign on the platform is that for most campaigns on the platform (including the one in our study), the advertiser pays only for ad exposures. Therefore, companies naturally adopt the ATT, or the ad effect (the "treatment effect") on a group of users (the "treated"), as the criterion for their business assessment. Second, in line with Johnson et al. (2017a), the definition of ATT delivers the correct strategic baseline because the potential outcome, Y_0, represents users' behavior if the focal advertiser does not run the campaign. Specifically, Y_0 involves the behavior conditional on what other ads the users would have seen had the focal advertiser not run the campaign (e.g., users could have seen ads from its competitor, which might have lowered their conversion for the focal brand). This is exactly the strategic baseline against which the advertiser wanted to measure its advertising effect. Finally, rather than measuring the effect of a particular ad creative, the ATT measures the effect of the whole advertising campaign, which is a collection of ad creatives. The main reason is that measuring the overall effect of the advertising campaign is of primary importance in the industry, including our industry partner. While measuring the effect of a single ad creative is important to optimize an advertising campaign, this issue is beyond the scope of our study.

4.3. Estimation and Inference

The estimation of ATT is based on Equation (5). We denote PL_i^{obs} as the observed preload status, W_i^{obs} as the observed ad exposure, and $\mathrm{Y}_i^{obs} \equiv \mathrm{Y}_i(\mathrm{Z}_i,\ PL_i^{obs}, \mathrm{W}_i^{obs})$ as the observed conversion outcome, for user i. We have

$$\hat{\pi} \equiv \widehat{ATT} = \frac{\widehat{ITT}_{PL}}{\hat{\rho}}, \tag{7}$$

where

$$\widehat{ITT}_{PL} = \hat{\mathrm{E}}\left[Y\left(1, PL(1), W\big(1, PL(1)\big)\right)\middle| PL(1) = 1\right] - \hat{\mathrm{E}}\left[Y\left(0, PL(0), W\big(0, PL(0)\big)\right)\middle| PL(1) = 1\right]. \tag{8}$$

The first item in Equation (8) can be estimated with the users in the treatment condition for whom the focal ad was preloaded. For the second item, because we do not observe users in both the treatment and control conditions, we estimate this item with users in the control condition who would be preloaded with the focal ad had they been assigned to the treatment group. Suppose that N_1 users in the treatment group were preloaded with the ad and N_0 users in the control group were supposed to be preloaded with the ad but were not because of experimental intervention. Then, we can estimate Equation (8) as

$$\widehat{ITT}_{PL} = \hat{\mathrm{E}}\left[Y\left(1, PL(1), W\big(1, PL(1)\big)\right)\middle| PL(1) = 1\right] - \hat{\mathrm{E}}\left[Y\left(0, PL(0), W\big(0, PL(0)\big)\right)\middle| PL(1) = 1\right]$$
$$= \frac{1}{N_1} \cdot \sum_{i=1}^{N_1} \mathrm{Y}_i^{obs} - \frac{1}{N_0} \cdot \sum_{i=1}^{N_0} \mathrm{Y}_i^{obs}. \tag{9}$$

We also have

$$\hat{\rho} = \frac{1}{N_1} \cdot \sum_{i=1}^{N_1} \mathbb{I}\big(\mathrm{W}_i^{obs} = 1\big). \tag{10}$$

Because users are randomized into treatment or control groups, $N_1^{-1} \sum_{i=1}^{N_1} \mathrm{Y}_i^{obs}$ is a consistent estimate of $\mathrm{E}\left[Y\left(1, PL(1), W\big(1, PL(1)\big)\right)\middle| PL(1) = 1\right]$, and $N_0^{-1} \sum_{i=1}^{N_0} \mathrm{Y}_i^{obs}$ is a consistent estimate of $\mathrm{E}\left[Y\left(0, PL(0), W\big(0, PL(0)\big)\right)\middle| PL(1) = 1\right]$. Therefore, $\widehat{ITT}_{PL}$ in Equation (9) is a consistent estimate of $\mathrm{ITT}_{\mathrm{PL}}$ defined in Equation (3). Similarly, $\hat{\rho}$ in Equation (10) is a consistent estimate of ρ. Therefore, the estimate of ATT, defined in Equation (7), is a consistent estimate of the advertising effect (i.e., the ATT) defined in Equation (2). Imbens and Angrist (1994) refer to the ATT as the local average treatment effect (LATE) and demonstrate that its estimate, as defined in Equation (7), is equivalent to the instrumental variable (IV) regression in which the independent variable is ad exposure or not and the IV is the treatment assignment. Similarly, the estimate for the ATT lift, $\hat{\tau}_l$, is

$$\hat{\pi}_l = \frac{\hat{\pi}}{\hat{E}\left[Y\left(1, PL(1), W\left(1, PL(1)\right)\right)\middle| W\left(1, PL(1)\right)=1, PL(1)=1\right]-\hat{\pi}}, \tag{11}$$

where

$$\hat{E}\left[Y\left(1, PL(1), W\left(1, PL(1)\right)\right)\middle| W\left(1, PL(1)\right) = 1, PL(1) = 1\right]$$

$$= \frac{1}{\sum_{i=1}^{N_1} \mathbb{I}(\mathrm{W}_i^{obs}=1)} \cdot \sum_{i=1}^{N_1} \mathrm{Y}_i^{obs} \cdot \mathbb{I}\left(\mathrm{W}_i^{obs} = 1\right). \tag{12}$$

For inference, we use the bootstrap method to calculate the confidence intervals for $\hat{\pi}$ and $\hat{\tau}_l$. Specifically, in bootstrap sample b, we randomly draw a sample of $N = N_1 + N_0$ users from the original sample with replacement. We estimate the ATT and ATT lift using this bootstrap sample, denoted as $\hat{\pi}^b$ and $\hat{\tau}_l^b$, respectively. We repeat this procedure B times to have a series of bootstrapped estimates of ATT and ATT lift, that is, $\left\{\hat{\pi}^b, \hat{\tau}_l^b\right\}_{b=1}^{B}$. We calculate the confidence intervals on the basis of this bootstrap sample.

5. Data

5.1. Estimation Sample

As discussed in section 4.3, the estimation of the advertising effect in our context requires a sample of users who were in the treatment group and preloaded with the focal ad, and also users who were in the control group and supposed to be preloaded with the focal ad but were not. The company provided us with this sample in accordance with our request. Hereinafter, the treatment and control groups refer only to those in our analysis sample.

The experiment covered more than 84 million users who were preloaded (in the treatment condition) or supposed to be preloaded (in the control condition) with the focal ads. The exposures were 44% on Douyin and 56% on Toutiao. For an exposed user, the average number of exposures was 2.2. Following the brand's common practice, we counted conversions as those that occurred at the start of the campaign (August 13, 2020) until half a month after the campaign ended (October 14, 2020). Hereinafter, we refer to this period (August 13, 2020 to October 14, 2020) as the conversion period.

In addition to the experiment data, the brand provided the complete list of conversions during the conversion period. For each conversion, also included in the data were consumers' cell phone numbers (encrypted to protect privacy) and the channel (within or outside ByteDance) on which the conversion occurred. The brand's data were matched with the experiment data using the encrypted cell phone numbers, which generated complete information about users' conversion behavior within and outside ByteDance. While all the online conversions were well tracked, the information cannot be acquired by the brand if a user directly visits a dealer. However, such incidences are rare according to the brand.

The final data contain each user's treatment status (treatment vs. control), exposure condition (exposed to the focal ad or not), and conversion outcome. If a user converts, the data records whether the conversion occurred on ByteDance or other places. In the data, the channels outside ByteDance are categorized into three types: the automobile-specified apps, the brand's own channels, and other digital media. ByteDance also provided key user demographics the brand used as its targeting criteria, including gender, city tiers (tiers 1 and 2 vs. tiers 3–5),[10] and age (≥31 vs. <31 years). In practice, these three demographics are also the most common variables ByteDance communicates to other automobile brands for targeting. The brand has its own enterprise account on Douyin. The data also include information on whether each user visited the brand's home page on Douyin within a month before the experiment (visit: 1; no visit: 0).

5.2. Summary Statistics

First, we checked the equivalence between the treatment and control groups in terms of demographic composition.[11] Following Huang et al. (2020), we conducted two one-sided t-tests (TOST) to investigate whether the difference between the two groups is sufficiently small. The tests, reported in Table 1, show

[10] The lower the tier, the more economically developed the city is.
[11] ByteDance's privacy concerns prevent us from reporting the distribution of the demographics for the treatment and control groups.

that the null hypotheses are rejected, and thus we conclude that the mean values of the demographics between the two groups are no greater than a small threshold.

Second, as shown in Table 2, we have roughly 42 million users each in the treatment and control groups. Among converted users, 75% converted just once, 22% converted twice, and the rest (3%) converted three times or more. The average number of conversions is similar between the treatment and control groups. For users with multiple conversions, 98.2% converted only in one channel (within or outside ByteDance), and only 1.8% converted from both within and outside ByteDance. In the table, “conversion within ByteDance” indicates whether a user ever converted from within ByteDance; if a user did this more than once, we treat the conversion as just one when calculating this number. The same principle applies to “conversion outside ByteDance.” “Overall conversion” indicates whether a user ever converted in the experiment. Therefore, we treat multiple conversions as just one conversion when calculating this number. This process explains why the sum of the conversions within and outside ByteDance sometimes exceeds the overall conversion.

Overall, the conversion rates in the treatment and control groups are 0.0348‰ and 0.0281‰, respectively, and the difference 0.0067‰ is significant ($p < 0.01$). For the control group, the conversion rate within ByteDance is zero due to no exposure, which represents the scenario that the brand does not advertise on ByteDance. However, 0.0281‰ of users in this group converted from outside ByteDance. In the treatment group, only 8.2% of users were exposed to the focal ads. The overall conversion rate of the exposed treatment users is roughly seven times that of unexposed treatment users. Notably, for the exposed treatment users, the conversion rate outside ByteDance is more than eight times that within ByteDance. Prior research has found similar spillover patterns; for example, the majority of the sales increase driven by online advertising happens in offline channels (Chan et al. 2011, Lewis and Reiley 2014). In both the treatment and control groups, the majority of the conversions outside ByteDance occurred in automobile-specialized apps (85.2% for treatment, 85.1% for control), and the rest are relatively equally shared by the brand’s own channels and the other digital media.

We also examine the demographics difference among group of users and report the results in Table A1–A3 in Online Appendix A. Table A1 shows that in the treatment group, the exposed users are more likely to be male, to come from tier 1 and 2 cities, to be older, and to have visited the brand's home page on Douyin within one month before the experiment than the unexposed users. Table A2 shows the same pattern for the exposed users in the treatment when compared with the control group. Finally, the different groups of converted users, including the control users converted outside ByteDance, the treatment users converted within ByteDance, and the treatment users converted outside ByteDance, are indifferent in terms of gender and age (see Table A3). However, treatment users converted within ByteDance are more likely to be from tier 1 and 2 cities. In addition, treatment users converted outside ByteDance are most likely to visit the brand's home page on Douyin within one month before the experiment, followed by control users converted outside ByteDance and treatment users converted within ByteDance.

6. Estimation

6.1. Advertising Effect and Economic Evaluation of Advertising Campaign

Table 3 shows the main estimation results. The ATT and lift for overall conversion are 0.0820‰ and 1.1226, respectively, and both are significant. This advertising effect is much greater than reported experiments on Google and Facebook. For example, Gordon et al. (2023) examined 663 randomized controlled trials with a mix of display and video ads run on Facebook in the United States in 2020. They found that the median lift on conversion in these experiments is only 9%. In addition, Johnson et al. (2017b) report a similar median lift in conversion (17% for site visits and 8% for conversions) in 432 experiments with display ads on Google Display Network.

This advertising effect is mainly driven by the conversions outside ByteDance. If we include only the conversions inside ByteDance, as brands commonly do, the ATT is 0.0171‰, which is statistically significant at the 5% level. As $\mathrm{E}\left[Y\left(0, PL(0), W\left(0, PL(0)\right)\right)\middle| W\left(1, PL(1)\right) = 1, PL(1) = 1\right] = 0$ for conversion within ByteDance (i.e., no conversion within ByteDance in the control condition), the associated ATT lift is ill-defined. However, simply for conversions outside ByteDance, the associated ATT is 0.0684‰ with a lift of 0.9362; both are significant at the 5% level.

These findings raise a serious question for brands conducting economic evaluations of short-form advertising on ByteDance. As mentioned previously, brands often complain about ByteDance's high price but low efficacy. In the Chinese advertising market, brands commonly use the CPC to economically evaluate their advertising campaigns. In current practice, for a campaign to run on an ad platform such as ByteDance, the brand calculates the CPC using only conversions within the platform, which we call a "naive partial CPC" and define as

$$CPC_{naivepartial} = \frac{AdExpense}{N \cdot r_{Platform}},$$

where $AdExpense$ is the total campaign expense, N is the number of users in the campaign, and $r_{Platform}$ is the conversion rate on the ad platform. If, however, the brand includes conversions outside the ad platform, the "naive complete CPC" is

$$CPC_{naivecomplete} = \frac{AdCost}{N \cdot r},$$

where r is the overall conversion rate. Thus, if we use the overall conversion rate and the conversion rate within ByteDance in Table 2, the ratio of $CPC_{naivecomplete}$ to $CPC_{naivepartial}$ is 0.0402; that is, incorporating the conversions outside ByteDance reduces the average CPC nearly 25 times. This reduction significantly lowers the quality-adjusted advertising price and increases the brand's economic evaluation of the advertising campaign.

However, this commonly adopted method is conceptually problematic, as the correct economic evaluation of the advertising campaign should be based on the treatment effect, such as ATT, which we label as $CPC_{ATTcomplete}$ and define as

$$CPC_{ATTcomplete} = \frac{AdCost}{N \cdot \rho \cdot ATT_{overall}},$$

where ρ, as defined previously, is the exposure ratio and $ATT_{overall}$ is the overall ATT. However, if the brand ignores the conversion outside the platform, it may calculate the CPC only on the basis of the ATT on the platform:

$$CPC_{ATTpartial} = \frac{AdCost}{N \cdot \rho \cdot ATT_{Platform}},$$

where $ATT_{Platform}$ is the ATT on the ad platform. Using the ATTs reported in Table 3, we find that the ratio of $CPC_{ATTcomplete}$ to $CPC_{ATTpartial}$ is 0.2085, indicating that inclusion of the conversions outside ByteDance shrinks the CPC to 20%.

6.2. Targeting Strategy in Advertising Campaign

On ByteDance, advertising campaigns for automobile brands are often large in scale, in an effort to reach a wide range of consumers. Therefore, instead of selecting many user features for targeting, which may result in coverage of only some niche groups, brands often rely on several key features in ByteDance's advertising system. As noted previously, gender, city tiers, and age are the most common targeting features ByteDance provides to automobile clients.

6.2.1. Heterogeneous Treatment Effects. We first analyze the advertising effect across these three variables. We present the estimation results of ATT and ATT lift for the segments by each demographic variable in Table 4. The advertising effects on conversion are all significantly positive for each segment. The conversion rate outside ByteDance is often several times that within ByteDance, which suggests that the overall effects are driven mainly by conversions outside ByteDance.

We then examine whether the treatment effects differ between segments within each demographic variable. For example, is advertising more effective for men or women? In line with Imbens and Rubin (2015), our estimator of ATT is equivalent to a two-stage least squares (2SLS) regression with conversion as the dependent variable, exposure as the independent variable, and the treatment assignment as the IV. Thus, we run the following 2SLS regression:

$$Y_i^{obs} = \alpha + \beta_0 \cdot W_i^{obs} + \beta_1 \cdot Male_i + \beta_2 \cdot Tier1or2City_i + \beta_3 \cdot AgeAbove30_i + \beta_{11} \cdot W_i^{obs} \cdot Male_i + \beta_{21} \cdot W_i^{obs} \cdot Tier1or2City_i + \beta_{31} \cdot W_i^{obs} \cdot AgeAbove30_i + \epsilon_i, \quad (13)$$

where i refers to the user, Y_i^{obs} refers to the conversion during the conversion period, W_i^{obs} denotes the observed ad exposure, $Male_i$ is a dummy variable for gender (male: 1; female: 0), $Tier1or2City_i$ is a dummy variable for the type of city (tiers 1 and 2: 1; otherwise: 0), and $AgeAbove30_i$ is a dummy variable for age (>30 years: 1; otherwise: 0). The IVs include Z_i, $Z_i \cdot Male_i$, $Z_i \cdot Tier1or2City_i$, and $Z_i \cdot AgeAbove30_i$, where Z_i refers to the treatment status.

We report the results in Table 5. Overall, ad exposure has no significant effect between segments within each demographic variable, which suggests that these typically used variables are ineffective for targeting. However, if we include only the conversion within ByteDance in the analysis, male and younger users appear more affected by the advertising, which could incorrectly lead the brand to allocate more budget to these segments. This finding again shows the importance of exchanging information between platforms and brands, which aids not only the economic evaluation of the advertising campaign but also the targeting strategy.

6.2.2. Behavioral Targeting. While we did not find any heterogeneous ad effects for the three most commonly used demographic variables, behavioral differences may moderate these effects. As such, we requested additional information from ByteDance on users' prior visits to the brand's home page on Douyin within one month before the experiment. The automobile brand considers home page visits an important measure of brand engagement. Such visits suggest that users are more familiar with and interested in the brand. We thus examine whether this behavior moderates the advertising effects on conversion.

For each user, we ran the following IV regression:

$$Y_i^{obs} = \alpha + \beta_0 \cdot W_i^{obs} + \beta_1 \cdot PriorVisit_i + \beta_3 \cdot W_i^{obs} \cdot PriorVisit_i + \epsilon_i, \quad (14)$$

where i refers to the user, Y_i^{obs} refers to conversion, W_i^{obs} denotes the observed ad exposure, and $PriorVisit_i$ is a dummy variable indicating whether the user visited the brand's home page within a month before the experiment (visit: 1; no visit: 0). The IVs include Z_i and $Z_i \cdot PriorVisit_i$, where Z_i refers to the treatment status.

We report the results in Table 6. As before, ad exposure increases users' likelihood of conversion. While prior visits to the brand's home page do not make a significant difference in terms of conversion likelihood, the advertising effect is greater for those with prior visits. When we include only conversions within ByteDance, we find no difference in the advertising effect between users with and without prior visits. Thus, a decision based on the conversion information within ByteDance will incorrectly lead the brand not to rely on the behavior information of prior home page visits.

A question is whether the effect of prior visits is robust after we include the demographic information. To address this, we run the following IV regression including demographic variables:

$$Y_i^{obs} = \alpha + \beta_0 \cdot W_i^{obs} + \beta_1 \cdot Male_i + \beta_2 \cdot Tier1or2City_i + \beta_3 \cdot AgeAbove30_i + \beta_4 \cdot PriorVisit_i + \beta_{11} \cdot W_i^{obs} \cdot Male_i + \beta_{21} \cdot W_i^{obs} \cdot Tier1or2City_i + \beta_{31} \cdot W_i^{obs} \cdot AgeAbove30_i + \beta_{41} \cdot W_i^{obs} \cdot PriorVisit_i + \epsilon_i, \quad (15)$$

where all the variables are defined the same as previously and the IVs include Z_i, $Z_i \cdot Male_i$, $Z_i \cdot Tier1or2City_i$, $Z_i \cdot AgeAbove30_i$, and $Z_i \cdot PriorVisit_i$. We report the results in Table 7, which shows that all the previous findings are robust. The advertising effects are homogeneous across the demographic variables but are moderated by prior visits to the brand's home page. Given the estimates, we calculate the conditional ATT of overall conversion for each combination of user features and report the results in Online Appendix B. The advertising effects on overall conversion are all significantly positive for each segment.

6.3. Further Analysis and Robustness Checks

6.3.1. Intertemporal Variation of Advertising Effect. Because the conversion period lasted for nearly two months from August to October, users' conversion could have been affected by incidences other than the advertising during that period. For example, in the China market, car dealers often run promotions in the fall, which could boost users' interest in the automobile product, consequently diluting the effects from advertising. If that is case, we would expect to find greater advertising effects earlier in the experiment.

Among converted users, 48.24% converted only during the campaign (stage 1 of the experiment), 48.92% converted only after the campaign (stage 2 of the experiment), and 2.83% converted during both periods. We calculate adverting's ATT on conversion during and after the campaign, respectively, and report the results in Table 8. The advertising effects are significant both during and after the campaign. Although the ATT and ATT lift are larger during the campaign than those after the campaign, the differences are not significant.

6.3.2. Advertising Effects by Conversion Channels. As mentioned previously, conversions could occur outside ByteDance in three types of channels: automobile-specialized apps, the brand's own channels, and other digital media. To unpack the effect of advertising among these channels, we calculate the ATT for conversions in each channel outside ByteDance and report the results in Table 9. The results show that advertising on ByteDance significantly increases the conversion rate in the automobile-specialized apps and the brand's own channels but not in the other digital media. While the advertising effect on brand's own channels is similar to that within ByteDance (Table 3), it is only 20% of that on the automobile-specialized apps. Although the ATT lift is greater on brand's own channels, it is not significantly different from that on automobile-specialized apps.

6.3.3. Moderating Effect of Number of Exposure. Because users could have been exposed multiple times in the experiment, we examine the moderating effect of number of exposures. Specifically, we run the following 2SLS regression:

$$Y_i^{obs} = \alpha + \beta_1 \cdot W_i^{obs} + \beta_2 \cdot W_i^{obs} \cdot I(NExposure_i \geq 2) + \beta_3 \cdot W_i^{obs} \cdot I(NExposure_i \geq 3) + \gamma_1 \cdot Male_i + \gamma_2 \cdot Tier1or2City_i + \gamma_3 \cdot AgeAbove30_i + \gamma_4 \cdot PriorVisit_i + \epsilon_i, \quad (16)$$

where $NExposure_i$ indicates the number of ad exposures for user i and all other variables are defined as before. In this case, parameter β_1 captures the advertising effect under only one exposure, β_2 is the incremental advertising effect from one to two exposures, and β_3 is the incremental ad effect from two to three or more exposures. The IVs include Z_i, $Z_i \cdot I(NExposure_i \geq 2)$, and $Z_i \cdot I(NExposure_i \geq 3)$, where Z_i refers to the treatment status. Note that conditional on exposure, the number of exposures is endogenously determined by the platform's algorithm. To mitigate such endogeneity, we include the demographics as controls because the algorithm partially relies on users' demographics. We report the results in Table 10. Overall, advertising has no effect on conversion if users are exposed only once. The effect significantly increases when users are exposed two or more times. While a similar pattern holds for conversions outside ByteDance, it differs for those within ByteDance, in which case one exposure significantly increases the conversion rate and the rate increases when three or more exposures occur. This finding is intuitive. Short-form video platforms are often at the top of the conversion funnel. Therefore, users converting on ByteDance are likely to be in the later stage of their decision process. Thus, one exposure is enough for conversion. By contrast, users exposed on ByteDance without conversion within ByteDance are likely to be in the early stage of the decision process. When they move to the later stage of the process and begin searching for information outside ByteDance, only additional exposures will have an effect, given memory decay of early exposed advertising.

7. Discussion

We examine the effect of an advertising campaign on ByteDance, a short-form video social media platform. In a large-scale randomized experiment, we show a significant effect of the advertising campaign, compared with effects reported in prior research. Of note, advertising spillover plays a pivot role in the advertising campaign. In our case, ignoring the conversion outside ByteDance would significantly underestimate the advertising effect and lead to an unfavorable economic evaluation of the advertising campaign. In addition, ignoring advertising spillover can mislead a brand's targeting strategy. We show that the demographic variables commonly used for targeting (i.e., gender, location, and age) do not work when considering advertising spillover; instead, behavioral variables such as prior brand home page visits may be more effective for targeting. However, if relying only on conversion within the ad platform, the data will incorrectly suggest that this behavioral variable does not work. Given that short-form video platforms are often located in the upper part of the conversion funnel, advertising on these platforms could naturally affect consumer behavior down the funnel. However, the scale of advertising spillover identified in our experiment is worth noting. As we show in the analysis, advertising spillover significantly affects the brand's cost evaluation and targeting strategy, which underscores the importance of information sharing between platforms and brands, which in practice is typically not the case.

Several points are worth highlighting. First, our estimates may not be an equilibrium outcome. During the experiment, the brand shifted all its digital advertising budget to ByteDance, and its competitors might not have been aware of this immediately. Although we do not have information on competitors' status, they might have discovered this shift and reacted at a certain point, given that the experiment lasted for nearly two months. If that were the case, we would have observed the advertising effect decreasing in the later stage of the experiment, given that competitors' reactions would most likely curb the brand's marketing effort. As Table 8 shows, we do find a smaller advertising effect later in the experimental period, though the difference is insignificant. Conceptually, estimating the equilibrium outcome using a structural approach is more convenient than using an experiment. Therefore, the focus of our study is not on deriving the equilibrium outcome but on identifying the large-scale advertising spillover, which could reasonably exist when the market reaches equilibrium.

A second important point is whether the experimental intervention shifted the distribution of ad quality for the control group. In the experiment, the focal ad is blocked before it is preloaded for users in the control group. As the focal ad is selected to be preloaded according to the algorithm for both open-screen and newsfeed advertising, a concern is whether blocking the focal ad systematically changed the quality distribution of the ads preloaded to users' cache, which would then have affected users' behavior in the control group. The following factors help mitigate this concern. In the experiment, when the focal ad was simply blocked (in open-screen ads) or replaced by another ad (in newsfeed ads), the rest of the ads in the batch remained the same. Given that 10 ads were often preloaded to the batch, the change in the quality distribution for ads preloaded in the batch was minor. In addition, even if a small change in the quality distribution occurred from the lack or replacement of the focal ad, this still represents the baseline case in the experiment (i.e., what will happen if the brand stops advertising on ByteDance?) (Johnson et al. 2017a). Finally, given that the algorithm, which is partially based on users' demographics, executes the blocking, we controlled for these user features when estimating the adverting effects. In Online Appendix B, we show that the conditional ATTs based on user segments are all significant.

A third important point is the heterogeneity over time and across ads. Given that the conversion period lasted for two months, the time effects could have shifted the likelihood of conversion. A hazard model approach could address this issue. However, the computation burden is enormous given the size of the data. In addition, we have no information on the timing of exposure, which makes this approach infeasible. In general, the impact of timing effects could be mild given that our focus is on whether a user converts or not within a certain period. With the randomization of the treatment and control groups, any shift in the likelihood of conversion due to timing effect would happen for both groups and therefore be cancelled out. For the heterogeneity across ads, we do not have information on the specific ads to which users were exposed. However, as the brand cares mostly about the overall effect of the campaign (not that of a single ad), this is not a major issue in the study.

Finally, a question is whether conversion is the only metric for the advertising campaign. Although the low conversion rate in the experiment is common in the automobile industry, people may wonder whether an advertising campaign is worthwhile in terms of the cost–benefit calculation. Automobile brands typically calculate the return of advertising campaigns on the basis of conversion (i.e., how many sales resulted from customer leads obtained from the advertising campaign). According to our conversations with industry experts, the average conversion rate for sales from customer leads is between 1.5% and 4%; that is, for each 1000 customer leads obtained, 15 to 40 cars are sold on average. With this information and the conversion rates in Table 2, we can estimate the number of cars sold through the campaign. Given the average car price and campaign expenditure,[12] the campaign likely broke even. However, if we calculate the return using the ATT in Table 3, the campaign is in a deficit. This suggests that other metrics, such as brand awareness and engagement, should be considered. Although we lack such measures for our experiment, future studies could include these metrics to unpack the full conversion rates.

[12] Our confidentiality agreement prohibits us from providing information on car price and campaign expenditure.

Funding and Competing Interests

One of the authors used to be an employee of Ocean Engine, which managed the experiment. All other authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed herein. The authors have no funding to report.

References

Aral S (2021) What digital advertising gets wrong. *Harvard Business Rev*. (February 19), https://hbr.org/2021/02/what-digital-advertising-gets-wrong.

Blake T, Nosko C, Tadelis S (2015) Consumer heterogeneity and paid search effectiveness: A large-scale field experiment. *Econometrica* 83(1):155-174.

Chan T, Wu C, Xie Y (2011) Measuring the lifetime value of customers acquired from google search advertising. *Marketing Sci.* 30(5):837–850.

Gordon BR, Jerath K, Katona Z, Narayanan S, Shin J, Wilbur KC (2021) Inefficiencies in digital advertising markets. *J. Marketing* 85(1):7-25.

Gordon BR, Moakler R, Zettelmeyer F (2023) Close enough? A large-scale exploration of non-experimental approaches to advertising measurement. *Marketing Sci.* 42(4):768-793.

Gordon BR, Zettelmeyer F, Bhargava N, Chapsky D (2019) A comparison of approaches to advertising measurement: Evidence from big field experiments at Facebook. *Marketing Sci.* 38(2):193–226.

Gutelle S (2022) TikTok's share of the U.S. influencer marketing industry is surpassing Facebook's. *Tubefilter* (August 2), https://www.tubefilter.com/2022/08/02/insider-intelligence-influencer-marketing-instagram-tiktok/.

Huang S, Aral S, Hu YJ, Brynjolfsson E (2020) Social advertising effectiveness across products: A large-scale field experiment. *Marketing Sci.* 39(6):1142-1165.

HubSpot (2023) Global social media trends report. https://offers.hubspot.com/social-media-trends-report.

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

Imbens G, Rubin D (2015) *Causal Inference for Statistics, Social, and Biomedical Sciences* (Cambridge University Press, New York).

Johnson G, Lewis R, Nubbemeyer E (2017a) Ghost ads: Improving the economics of measuring online ad effectiveness. *J. Marketing Res.* 54(6):867-884.

Johnson G, Lewis R, Nubbemeyer E (2017b) The online display ad effectiveness funnel & carryover: Lessons from 432 field experiments. Working paper, Questrom School of Business.

Lewis R, Reiley D (2014) Online ads and offline sales: Measuring the effects of retail advertising via a controlled experiment on Yahoo! *Quantitative Marketing and Econ.* 12:235–266.

Rosenfeld B (2022) How marketers are fighting rising ad costs. *Forbes* (November 14), https://www.forbes.com/sites/forbescommunicationscouncil/2022/11/14/how-marketers-are-fighting-rising-ad-costs/?sh=5933c0622829.

Shore V (2023) Evolution of short form video marketing. Storyblocks (July 24), https://blog.storyblocks.com/marketing/evolution-of-short-form-video-marketing/.

Terlep S, Vranica S, Raice S (2012) GM says Facebook ads don't pay off. *The Wall Street Journal* (May 16), https://www.wsj.com/articles/SB10001424052702304192704577406394017764460.

Vidyard (2023) Video in business benchmark report. https://www.vidyard.com/business-video-benchmarks/.

Vranica S (2018) P&G Contends too much digital ad spending is a waste. *The Wall Street Journal* (March 1). https://www.wsj.com/articles/p-g-slashed-digital-ad-spending-by-another-100-million-1519915621.

Winter D (2023) TikTok ad revenue: How much are brands spending on TikTok? Shopify (July 14), https://www.shopify.com/blog/tiktok-ad-spending.

Yang J, Zhang J, Zhang Y (2021) First law of motion: Influencer video advertising on TikTok. Working paper, Harvard University.

Table 1. Balance in Demographics

	t-stat for TOST	
	H_0: $\mu_c - \mu_t < -\Delta$	H_0: $\mu_c - \mu_t > \Delta$
Male	20.4909***	-30.4542***
City tier 1 or 2	20.7221***	-20.3276***
Age ≥ 31	25.7473***	-41.7289***
Prior visit to brand's Douyin page	4.0805***	-13.9304***

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. *Notes*. The variable "Prior visit to brand's Douyin page" is an indicator of whether the user visited the brand's home page on Douyin within one month before the experiment. We ran an equivalence test H_0: $|\mu_c - \mu_t| > \Delta$, where $\Delta = 0.5\% \times \overline{X}_{treatment}$ (Huang et al. 2020), which is two one-sided t-tests (TOST. The H_0 of the first t-test is $\mu_c - \mu_t < -\Delta$ against H_1: $\mu_c - \mu_t \geq -\Delta$, while the second H_0 is $\mu_c - \mu_t > \Delta$ against H_1: $\mu_c - \mu_t \leq \Delta$. The tests conclude that $|\mu_c - \mu_t| \leq \Delta$.

Table 2. Summary Statistics

	Control	Treatment		
		Total	Exposed	Unexposed
No. users	42,135,109	42,136,733	3,451,204	38,685,529
Exposure rate	------	0.0819	------	------
Conversion (‱)				
Overall	0.0281	0.0348	0.1550	0.0240
Within ByteDance	------	0.0014	0.0171	------
Outside ByteDance	0.0281	0.0337	0.1414	0.0240
Automobile-specialized apps	0.0239	0.0287	0.1205	0.0205
Brand's own channels	0.0023	0.0032	0.0154	0.0021
Other digital media	0.0022	0.0023	0.0075	0.0018
Mean N. of conversions among converted users	1.2775	1.2908	1.3065	1.2817

Table 3. ATT and Lift

	ATT (‱)	ATT lift
Conversion:		
Overall	0.0820* [0.0545, 0.1121]	1.1226* [0.5702, 2.3761]
Within ByteDance	0.0171* [0.0130, 0.0217]	------
Outside ByteDance	0.0684* [0.0409, 0.0980]	0.9362* [0.4344, 2.0903]

* Significant at the 5% level. *Notes*. The bias-corrected bootstrap 95% confidence intervals are in brackets (B = 10000).

Table 4. Heterogeneous Treatment Effects

	Gender		City		Age	
	Male	Female	Tiers 1 and 2	Tiers 3–5	≤30	≥31
Conversion: ATT (‰)						
Overall	0.0822* [0.0432, 0.1220]	0.0814* [0.0447, 0.1149]	0.0845* [0.0497, 0.1206]	0.0790* [0.0295, 0.1272]	0.1022* [0.0385, 0.1670]	0.0740* [0.0423, 0.1062]
Within ByteDance	0.0196* [0.0140, 0.0257]	0.0130* [0.0076, 0.0199]	0.0179* [0.0127, 0.0248]	0.0161* [0.0103, 0.0225]	0.0222* [0.0141, 0.0324]	0.0150* [0.0106, 0.0203]
Outside ByteDance	0.0658* [0.0267, 0.1051]	0.0722* [0.0362, 0.1043]	0.0702* [0.0359, 0.1056]	0.0661* [0.0172, 0.1142]	0.0850* [0.0216, 0.1487]	0.0618* [0.0301, 0.0935]
Conversion: ATT lift						
Overall	0.8048* [0.3209, 1.7939]	3.1660* [0.7982, 77.9388]	1.4160* [0.5600, 4.2107]	0.8842* [0.2281, 2.8022]	1.0344* [0.2511, 3.9438]	1.1836* [0.4733, 2.9642]
Within ByteDance	------	------	------	------	------	------
Outside ByteDance	0.6449* [0.1993, 1.5425]	2.8089* [0.6420, 70.0593]	1.1771* [0.4039, 3.6930]	0.7404* [0.1343, 2.4950]	0.8606* [0.1419, 3.5451]	0.9887* [0.3424, 2.6191]

* Significant at the 5% level. *Notes*. The bias-corrected bootstrap 95% confidence intervals are in brackets (B = 10000).

Table 5. Heterogeneous Treatment Effects: 2SLS Regression

	Conv. overall	Conv. within ByteDance	Conv. outside ByteDance
Intercept	0.0264*** (0.0020)	------	0.0264*** (0.0020)
Dummy (male)	0.0184*** (0.0017)	------	0.0184*** (0.0017)
Dummy (city tiers 1 and 2)	-0.0031* (0.0017)	------	-0.0031* (0.0017)
Dummy (age ≥ 31)	-0.0105*** (0.0019)	------	-0.0105*** (0.0018)
AdExposure	0.0991** (0.0422)	0.0172*** (0.0045)	0.0871** (0.0418)
AdExposure × dummy (male)	0.0004 (0.0318)	0.0065* (0.0034)	-0.0067 (0.0315)
AdExposure × dummy (city tiers 1 and 2)	0.0049 (0.0306)	0.0016 (0.0032)	0.0037 (0.0303)
AdExposure × dummy (age ≥ 31)	-0.0282 (0.0340)	-0.0070** (0.0036)	-0.0233 (0.0337)
N	84,271,842	84,271,842	84,271,842
R^2	0.000022	0.000018	0.000018

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. *Notes.* The IV is the indicator for the treatment group. All estimates and SEs are multiplied by 1000, so that their interpretation is consistent with the unit of ‰ for the ATTs in the previous tables.

Table 6. Advertising Effect on Conversion by Prior Visit

	Dummy dependent variables		
	Conv. overall	Conv. within ByteDance	Conv. outside ByteDance
Intercept	0.0281*** (0.0009)	------	0.0281*** (0.0009)
PriorVisit	0.0005 (0.0134)	------	0.0005 (0.0133)
AdExposure	0.0780*** (0.0149)	0.0172*** (0.0016)	0.0643*** (0.0148)
AdExposure × PriorVisit	0.8376*** (0.2067)	-0.0172 (0.0217)	0.8513*** (0.2048)
N	84,271,842	84,271,842	84,271,842
R^2	0.000018	0.000016	0.000015

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. *Notes.* The IV is the indicator for the treatment group. All estimates and SEs are multiplied by 1000.

Table 7. Advertising Effect on Conversion by Prior Visit: Control for Demographics

	Dummy dependent variables		
	Conv. overall	Conv. within ByteDance	Conv. outside ByteDance
Intercept	0.0264*** (0.0020)	------	0.0264*** (0.0020)
Male	0.0184*** (0.0017)	------	0.0184*** (0.0017)
City Tiers 1 and 2	-0.0031* (0.0017)	------	-0.0031* (0.0017)
Age ≥ 31	-0.0105*** (0.0019)	------	-0.0105*** (0.0018)
PriorVisit	-0.0003 (0.0134)	------	-0.0003 (0.0133)
AdExposure	0.0943** (0.0422)	0.0173*** (0.0045)	0.0821** (0.0419)
AdExposure × male	0.0004 (0.0318)	0.0065* (0.0034)	-0.0067 (0.0315)
AdExposure × (city tiers 1 and 2)	0.0047 (0.0306)	0.0016 (0.0032)	0.0035 (0.0303)
AdExposure × (age ≥ 31)	-0.0267 (0.0340)	-0.0071** (0.0036)	-0.0219 (0.0337)
AdExposure × PriorVisit	0.8357*** (0.2067)	-0.0177 (0.0217)	0.8498*** (0.2048)
N	84,271,842	84,271,842	84,271,842
R^2	0.000023	0.000018	0.000018

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. *Notes.* The IV is the indicator for the treatment group. All estimates and SEs are multiplied by 1000.

Table 8. ATT and Lift for During- and After-Campaign Conversions

	ATT (‰)	ATT lift
Conversion:		
Overall	0.0820* [0.0545, 0.1121]	1.1226* [0.5702, 2.3761]
During campaign	0.0504* [0.0302, 0.0726]	1.4867* [0.6101, 4.7911]
After campaign	0.0342* [0.0142, 0.0554]	0.8486* [0.2558, 2.5648]

* Significant at the 5% level. *Notes*. The bias-corrected bootstrap 95% confidence intervals are in brackets (B = 10000).

Table 9. ATT and Lift for Outside-ByteDance Channels

	ATT (‰)	ATT lift
Conversion:		
Automobile-specialized apps	0.0588* [0.0322, 0.0856]	0.9527* [0.3845, 2.2547]
Brand's own channels	0.0107* [0.0035, 0.0177]	2.3117* [0.3171, 66.8986]
Other digital media	0.0012 [-0.0064, 0.0078]	0.1816 [-0.5003, 11.3925]

* Significant at the 5% level. *Notes*. The bias-corrected bootstrap 95% confidence intervals are in brackets (B = 10000).

Table 10. Moderation Effect of Number of Exposure

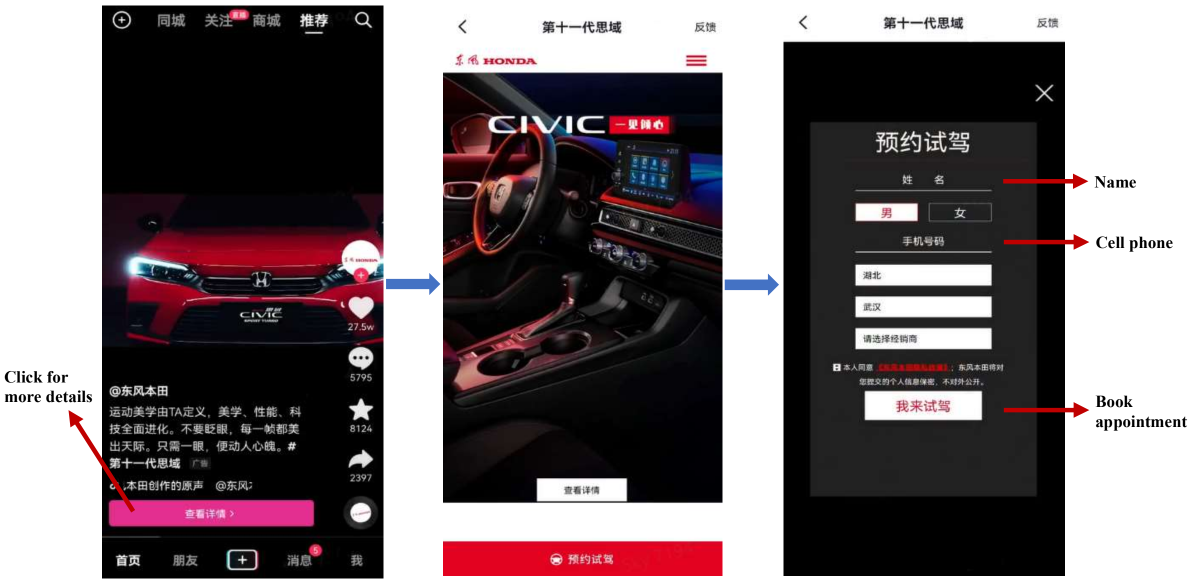
	Dummy dependent variables		
	Conv. overall	Conv. within ByteDance	Conv. outside ByteDance
Intercept	0.0287*** (0.0015)	------	0.0286*** (0.0015)
Male	0.0170*** (0.0012)	------	0.0168*** (0.0012)
City Tiers 1 and 2	-0.0040*** (0.0012)	------	-0.0040*** (0.0012)
Age ≥ 31	-0.0123*** (0.0013)	------	-0.0120*** (0.0013)
PriorVisit	0.0381*** (0.0095)	------	0.0387*** (0.0094)
AdExposure	0.0134 (0.0217)	0.0107*** (0.0023)	0.0048 (0.0215)
AdExposure × I(N. Exposure ≥ 2)	0.1385*** (0.0225)	-0.0003 (0.0026)	0.1367*** (0.0223)
AdExposure × I(N. Exposure ≥ 3)	0.1252*** (0.0107)	0.0346*** (0.0016)	0.1015*** (0.0106)
N	84,271,842	84,271,842	84,271,842
R^2	0.000028	0.000027	0.000023

****p* < 0.01, ***p* < 0.05, **p* < 0.1. *Notes.* The IV is the indicator for the treatment group. All estimates and SEs are multiplied by 1000.

Figure 1. Example of Ads on ByteDance

a. Open-Screen Ad on Douyin

b. Newsfeed Ad on Douyin

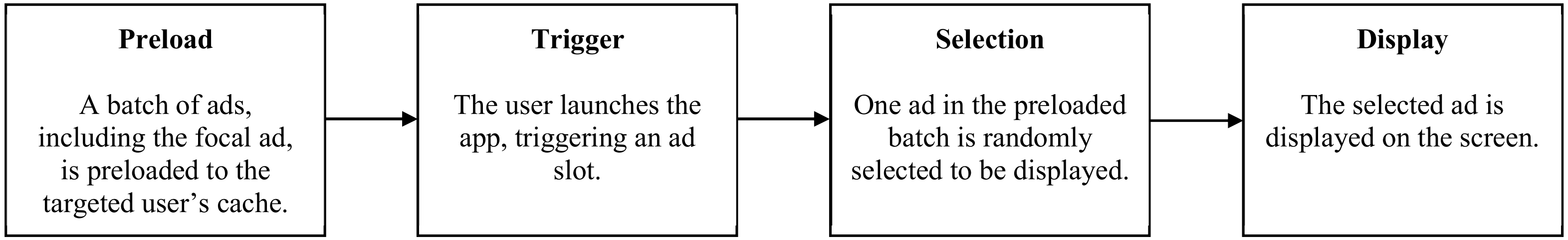


c. Newsfeed Ad on Toutiao

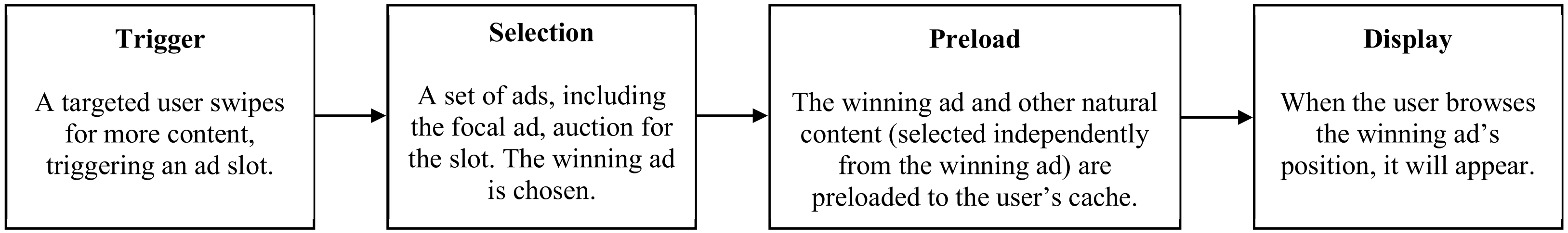
Figure 2. Experiment Procedure

a. Treatment Group

Open-screen Ads



Newsfeed Ads



b. Control Group

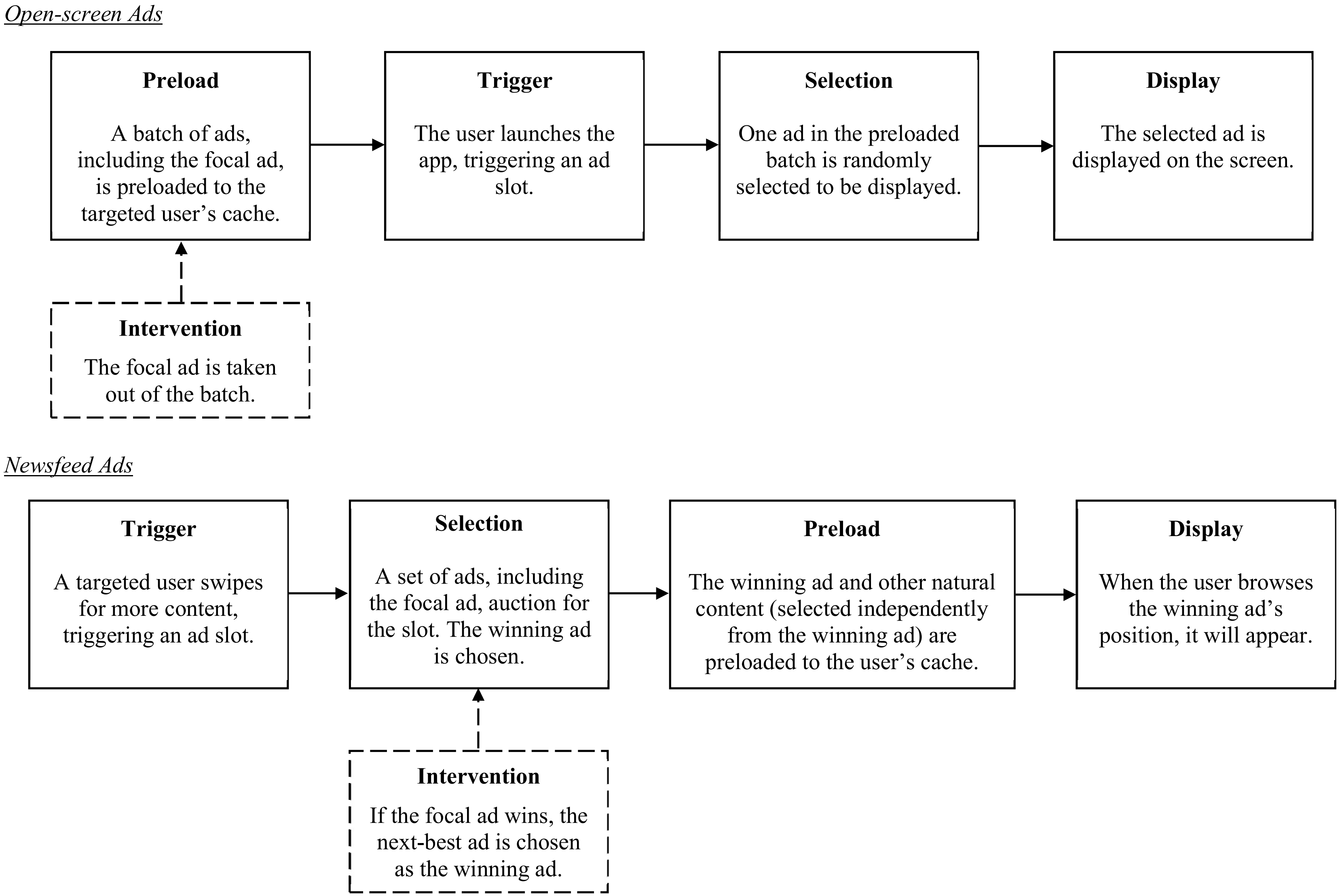
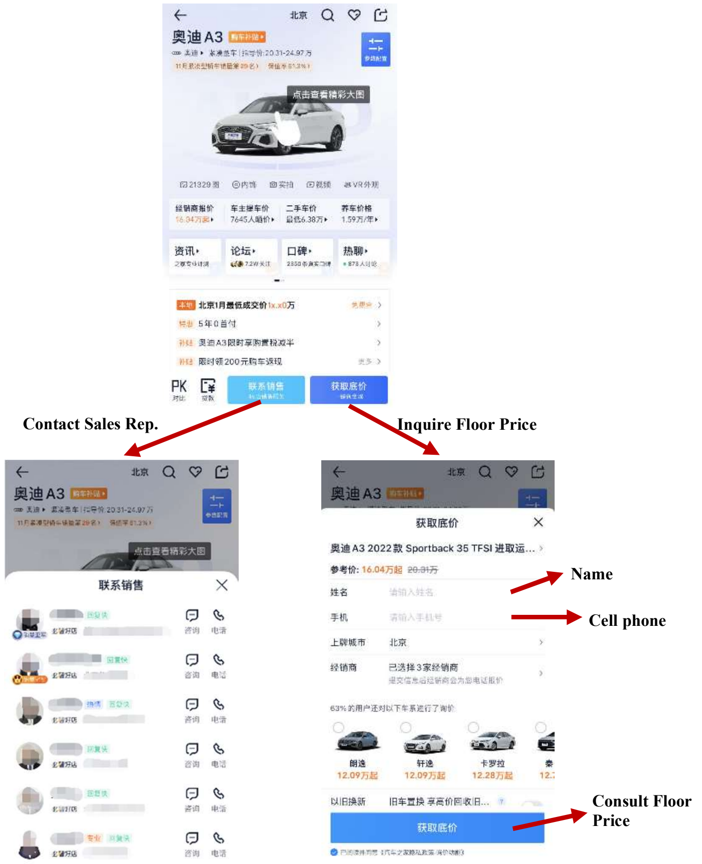


Figure 3. Example of Conversion in Autohome (Outside ByteDance)



Online Appendix A. Demographic Difference among Groups of Users

Table A1. Demographic Comparison between Exposed and Unexposed Users in the Treatment Group

	t-stat for TOST	
	H_0: $\mu_{unexpose} - \mu_{expose} < -\Delta$	H_0: $\mu_{unexpose} - \mu_{expose} > \Delta$
Male	-264.0934	-286.8319***
City tier 1 or 2	-397.4092	-417.0496***
Age ≥ 31	-111.5573	-139.5547***
Prior visit to brand's Douyin page	-10.4064	-17.7962***

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. *Notes*. The variable "Prior visit to brand's Douyin page" is an indicator of whether the user visited the brand's home page on Douyin within one month before the experiment. We ran equivalence test H_0: $|\mu_{unexpose} - \mu_{expose}| > \Delta$, where $\Delta = 0.5\% \times \overline{X}_{expose}$ (Huang et al. 2020), which is two one-sided t-tests (TOST). The H_0 of the first t-test is $\mu_{unexpose} - \mu_{expose} < -\Delta$ against H_1: $\mu_{unexpose} - \mu_{expose} \geq -\Delta$, while the second H_0 is $\mu_{unexpose} - \mu_{expose} > \Delta$ against H_1: $\mu_{unexpose} - \mu_{expose} \leq \Delta$. The tests conclude that $\mu_{unexpose} - \mu_{expose} < -\Delta$.

Table A2. Demographic Comparison between Exposed Treated Users and Control Users

	t-stat for TOST	
	H_0: $\mu_{control} - \mu_{expose} < -\Delta$	H_0: $\mu_{control} - \mu_{expose} > \Delta$
Male	-244.3900	-267.2109***
City tier 1 or 2	-365.1340	-384.8367***
Age ≥ 31	-104.8373	-132.9378***
Prior visit to brand's Douyin page	-11.0952	-18.5088***

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. *Notes*. The variable "Prior visit to brand's Douyin page" is an indicator of whether the user visited the brand's home page on Douyin within one month before the experiment. We ran equivalence test H_0: $|\mu_{control} - \mu_{expose}| > \Delta$, where $\Delta = 0.5\% \times \overline{X}_{expose}$ (Huang et al. 2020), which is two one-sided t-tests (TOST). The H_0 of the first t-test is $\mu_{control} - \mu_{expose} < -\Delta$ against H_1: $\mu_{control} - \mu_{expose} \geq -\Delta$, while the second H_0 is $\mu_{control} - \mu_{expose} > \Delta$ against H_1: $\mu_{control} - \mu_{expose} \leq \Delta$. The tests conclude that $\mu_{control} - \mu_{expose} < -\Delta$.

Table A3. Demographic Comparison among Converted Users

	t-stat (p value) for control vs. treatment within ByteDance: H_0: $\mu_0 - \mu_1 = 0$		t-stat (p value) for control vs. treatment outside ByteDance: H_0: $\mu_0 - \mu_2 = 0$		t-stat (p value) for treatment within vs. outside ByteDance: H_0: $\mu_1 - \mu_2 = 0$	
	H_1: $\mu_0 - \mu_1 > 0$	H_1: $\mu_0 - \mu_1 < 0$	H_1: $\mu_0 - \mu_2 > 0$	H_1: $\mu_0 - \mu_2 < 0$	H_1: $\mu_1 - \mu_2 > 0$	H_1: $\mu_1 - \mu_2 < 0$
Male	-0.4797 (0.6832)	-0.4797 (0.3168)	0.9783 (0.1640)	0.9783 (0.8360)	0.7500 (0.2284)	0.7500 (0.7716)
City tier 1 or 2	-1.9813 (0.9734)	-1.9813** (0.0266)	-1.1007 (0.8644)	-1.1007 (0.1356)	1.6961** (0.0481)	1.6961 (0.9519)
Age ≥ 31	-0.4972 (0.6894)	-0.4972 (0.3106)	-0.4081 (0.6584)	-0.4081 (0.3416)	0.3895 (0.3493)	0.3895 (0.6507)
Prior visit to brand's Douyin page	2.2399** (0.0126)	2.2399 (0.9874)	-2.7155 (0.9967)	-2.7155*** (0.0033)	-4.5027 (1.0000)	-4.5027*** (0.0000)

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. *Notes*. The variable "Prior visit to brand's Douyin page" is an indicator of whether the user visited the brand's home page on Douyin within one month before the experiment.

Online Appendix B. Conditional ATT of Overall Conversion

Combination of user features				ATT
Male	City tiers 1–2	Age ≥ 31	Prior visit	0.9083*** (0.2071)
Male	City tiers 1–2	Age ≥ 31	No prior visit	0.0726*** (0.0229)
Male	City tiers 1–2	Age < 31	Prior visit	0.9350*** (0.2078)
Male	City tiers 1–2	Age < 31	No prior visit	0.0993*** (0.0321)
Male	City tiers 3–5	Age ≥ 31	Prior visit	0.9036*** (0.2076)
Male	City tiers 3–5	Age ≥ 31	No prior visit	0.0679** (0.0278)
Male	City tiers 3–5	Age < 31	Prior visit	0.9303*** (0.2084)
Male	City tiers 3–5	Age < 31	No prior visit	0.0947*** (0.0365)
Female	City tiers 1–2	Age ≥ 31	Prior visit	0.9079*** (0.2078)
Female	City tiers 1–2	Age ≥ 31	No prior visit	0.0722** (0.0292)
Female	City tiers 1–2	Age < 31	Prior visit	0.9346*** (0.2087)
Female	City tiers 1–2	Age < 31	No prior visit	0.0989*** (0.0380)
Female	City tiers 3–5	Age ≥ 31	Prior visit	0.9032*** (0.2084)
Female	City tiers 3–5	Age ≥ 31	No prior visit	0.0675** (0.0338)
Female	City tiers 3–5	Age < 31	Prior visit	0.9299*** (0.2094)
Female	City tiers 3–5	Age < 31	No prior visit	0.0943** (0.0422)

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. *Notes.* All estimates and SEs are multiplied by 1000, so that their interpretation is consistent with the unit of ‰ for the ATTs in the previous tables.