# AdverTiming Matters: Examining User Ad Consumption for Effective Ad Allocations on Social Media


KOUSTUV SAHA, Georgia Institute of Technology, USA

YOZEN LIU, Snap Inc., USA

NICHOLAS VINCENT, Northwestern University, USA

FARHAN ASIF CHOWDHURY, University of New Mexico, USA

LEONARDO NEVES, Snap Inc., USA

NEIL SHAH, Snap Inc., USA

MAARTEN W. BOS, Snap Inc., USA



Showing ads delivers revenue for online content distributors, but ad exposure can compromise user experience and cause user fatigue and frustration. Correctly balancing ads with other content is imperative. Currently, ad allocation relies primarily on demographics and inferred user interests, which are treated as static features and can be privacy-intrusive. This paper uses person-centric and momentary context features to understand optimal ad-timing. In a quasi-experimental study on a three-month longitudinal dataset of 100K Snapchat users, we find ad timing influences ad effectiveness. We draw insights on the relationship between ad effectiveness and momentary behaviors such as duration, interactivity, and interaction diversity. We simulate ad reallocation, finding that our study-driven insights lead to greater value for the platform. This work advances our understanding of ad consumption and bears implications for designing responsible ad allocation systems, improving both user and platform outcomes. We discuss privacy-preserving components and ethical implications of our work.



CCS Concepts: • **Human-centered computing** → *Empirical studies in collaborative and social computing*; *Social media*; • **Applied computing** → *Psychology*.

Additional Key Words and Phrases: social media, ads, online ad, Snapchat, momentary behavior, causal-inference

**ACM Reference Format:**

Koustuv Saha, Yozen Liu, Nicholas Vincent, Farhan Asif Chowdhury, Leonardo Neves, Neil Shah, and Maarten W. Bos. 2020. AdverTiming Matters: Examining User Ad Consumption for Effective Ad Allocations on Social Media. In *Proceedings of* . ACM, New York, NY, USA, 25 pages. https://doi.org/10.1145/1122445.1122456


## 1 INTRODUCTION

In recent years, many online platforms predominantly generate revenue from advertisements (ads). Ad revenue offsets costs, making the services "free" to use — ad-supported business models are even considered to be at the "heart of the free internet" [100]. Some common examples are search engine services like Google and Bing, and social platforms like Facebook, StackExchange, LinkedIn, and Snapchat. These online platforms show ads via a diversity of implicit and explicit mechanisms such as "sponsored" or "promoted" content. However, an online platform that relies on ad revenue


Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2020 Association for Computing Machinery.

Manuscript submitted to ACM

must contend with the tradeoff between ad revenue and ads' impact on users. If a platform shows more ads, it runs the risk of hurting user experience and losing userbase [10, 12, 16]. Consequently, platforms resort to optimizing ad allocations that aim for multi-stakeholder benefits from user-centric, platform-centric, and advertiser-centric perspectives.

Traditionally, ad delivery in mass media such as print and television took a blanket approach — the same ad was shown to everyone who read the same newspaper or watched the same television channel at a time, and accordingly, demography (age/gender)-based targeting was conducted using people's interests (e.g. certain ads might only play on a sports channel). Although a blanket approach to advertising somewhat works in new and online media, online platforms introduce new complexities, potentials and dimensions [10]. With the ubiquity of various personalizations, content and ad delivery is often algorithmically customized to suit the interests of a specific user and improve their engagement with the platform. Importantly, the objective of personalized ad allocation is not just to increase revenue per user (and ad), but also to improve a user's satisfaction and experience with ads — a user would plausibly be more interested in ads topically relevant to their interests [60]. Personalized ad allocation has therefore built on the success of research in Recommender Systems, Machine Learning, and Human-Computer-Interaction (HCI) [20, 35, 45]. However, content personalization algorithms typically rely on demographic attributes like age, race, and gender, which are not only privacy-intrusive but are also static and exclusionary. As such, this practice has been critiqued in the fairness, accountability, transparency literature as reinforcing (and potentially exacerbating) stereotypes and societal biases [3, 40, 81].

While demographic attributes and content-based recommendations have been tremendously explored, other factors remain relatively less known — online ad allocation and ad spacing strategies typically rely on sets of rules, such as ads are shown after T duration or after N content views, where the parameters are mostly drawn from observations from the average user on a platform. However, the same ad allocation strategy would not necessarily work effectively on all users, given that every individual is different, and that they have a different lifestyle, behavior, needs, and engagement both offline and online [6]. In fact, online behaviors are also functions of offline context and routines, as well as users' momentary psychological and cognitive states [36, 63]. Therefore, it is important to embrace and evaluate dynamic and context- centric ad allocation strategies. This line of work remains underexplored both generally and particularly on evolving and newer forms of social and online content delivery.

Our work aims to address the above gap by studying ad consumption on Snapchat, a popular multimedia-driven online social platform which has a content feed called "Discover Feed", similar to content feeds offered by Facebook, Twitter, and Instagram's News or Story feeds. We latch on to the notion that users may show varying ad-consumption behaviors by the time of the day or other in-the-moment activities on the platform. Theoretically, our work is motivated by the body of literature that explains the psychological and time of the day effects in content and ad consumption [35, 98]. Practically, our work builds on the motivation that by teasing out these effects in a person-centric and context-centric fashion, we can not only draw better insights about ad consumption but can also make specific recommendations regarding *when* and *whom* questions in "What be shown to whom, and when?" — a question that interests both content providers (platform owners) as well as consumers (users). In particular, our work proposes three *research aims*:

**Aim 1:** To examine the effect of showing ads at a time preferred by users.
**Aim 2:** To examine how ad effectiveness varies with in-the-moment user behaviors on the platform.
**Aim 3:** To estimate the effect of ad allocations based on insights drawn from the above two aims.

We conduct our work on three-month longitudinal data of 100K Snapchat users. First, we conduct a quasi-experimental study to find that timing ads plays a strong effect in increasing ad effectiveness, as measured by ad reception and ad click-through rate (CTR). Then, we examine the relationship of ad effectiveness with momentary on-platform behaviors. Specifically, we consider how various measures — duration, activity, interactivity, interaction diversity, distractedness, and extra-socialness — might reflect insights about when to conduct ad allocations with respect to these measures. Finally, we estimate the potential impact of our insights by simulating experiments of reallocating ads. We find that simulations adopting insight about the relationship between ad reception and various patterns in on-platform behavior from our study guide more balanced and effective distributions of ads.

From a methodological perspective, this work contributes a causal-framework of modeling and inferring user ad consumption, which can extend to study other forms of user engagement using observational social media data. The efficacy of our operationalized treatment measure of *preferred ad timings* reveals the potential to use this novel measure in ad and other content engagement mechanisms. From an HCI perspective, our work augments the rich body of work studying the need to balance user experience and ad effectiveness [4, 55, 87, 102]. Theoretically, our work bears implications in advancing our knowledge of ad consumption on social media. We discuss the implications for designing better and more responsible ad allocation systems from a multi-stakeholder perspectives. By taking a broader view of ad allocation, we argue that it is possible to create better outcomes for both users and platforms. As an implication, our ad allocation approach's efficacy would help platforms to accomplish monetary goals with fewer ads, and therefore, can lead to allocating fewer ads — a solution that would be appreciated by users. Finally, a key advantage of our computational framework lies in the fact that it does not use demographic, typographic, or privacy-sensitive information of the users, we also discuss the privacy-preserving component and ethical implications of this work.

**Privacy, Ethics, and Disclosure.** This paper uses sourced data on Snapchat. Our work is conducted within Snapchat, and given the sensitivity of our work, we are committed to securing the privacy of the individuals. The dataset was accessed within a secured environment with necessary privacy and ethical protocols in place. The dataset was de-identified and no personally identifiable or demographic information was used. This paper only reports aggregated and $z$-transformed measures to prevent traceability and identifiability of users, and to prevent disclosing company-private information, yet providing context in readership. Even accounting for the benefits, we recognize the potential misuses, risks, and ethical consequences involved with this kind of research, on which we elaborate in the Discussion.

## 2 BACKGROUND AND RELATED WORK

### 2.1 Ad Effectiveness and User Experience

Conceptually, ad effectiveness is a key indicator of the success of an ad based on how well it does or the returns it generates in various forms across user likeability, engagement, and sales [101]. In an early work, Morrison and Dainoff studied ad complexity and dwell times, i.e., how much time a user spends to look at an ad and if they remember an ad more than others [71]; dwell times have been widely used as an implicit metric to study user interest and satisfaction [44, 54]. Doyle and Saunders defined effective ads as those that help advertisers reach their goals [21]. Ducoffe developed survey scales to measure ad effectiveness in terms of ad value in traditional media [22], which was later extended in the online media [23], positing ad value as a form of *communication engagement* between advertisers and consumers [23, 24]. Ducoffe and Curlo followed up to propose quantifiable concepts of expected advertising value (EAV) and outcome advertising value (OAV) of ads [24]. These assessments have also been used in-practice and in comparing online and offline ad effectiveness. In the online form, ad effectiveness is often quantified as return rate or Click-Through-Rate (CTR) [86]. CTR essentially measures the proportion of effectively allocated ads or the ratio of the clicks on an ad to its number of impressions [14]. Other work has proposed sophisticated measurements of online ad effectiveness such as that using ghost ads and experimental approaches [42].

Ad effectiveness is considered a vital outcome while planning, creating, and executing an ad [82]. Research has studied how various factors relate with ad value [24, 62, 83]. These studies found that informativeness and entertainment aspect bear a positive association with ad effectiveness, whereas intrusiveness bears a negative relationship [23]. Further, ad effectiveness shares a deep and complex interplay with user experience on the platform [10, 72]. Brajnik and Gabrielli reviewed the effects of online advertising on user experience and proposed a systematic theoretical framework for its better understanding [10]. Ads can cause fatigue, irritation, and negative emotions on users, making them leave and reduce engagement on the platform, consequently hurting both ad effectiveness and platform engagement [10]. Therefore, it is critical to optimize ad allocation in such a way that user experience is not compromised, as shown in recent HCI research through gamification [4], intelligent placement strategies [72], and animation [20].

Our work draws motivation from the above body of work to operationalize and study factors associated with ad effectiveness on social media. We extend the HCI community's long-drawn interest in balancing user experience with ad effectiveness [4, 16, 20, 72, 87, 102]. We define ad effectiveness using two measures based on 1) what fraction of time a user fully watches an ad (or ad reception), and 2) whether a user expresses some form of interest in the ad by clicking on it (or ad CTR). We then examine the role of (previously unexplored) factors such as timing and momentary on-platform behavior in explaining ad effectiveness outcomes in online platforms.

### 2.2 Ad Consumption Contextualized with Psychological Factors

Marketing and consumer research has extensively studied the importance of "antecedent state" — a term that encompasses all of the momentary financial, psychological, and physiological attributes with which a consumer arrives at a marketing interaction [8]. Haugtvedt et al. studied how personality traits associate with ad effectiveness [38]. In particular, mood states are known to significantly influence consumer behavior, judgment, and recall [28], and within the space of online ads, beliefs and attitude towards ads have been identified to predict ad effectiveness [25, 107]. Batra and Stayman showed that positive mood mediates brand attitudes in print ads [7], and Edwards et al. adopted the lens of psychological reactance to understand forced responses to ads and correspondingly the perceived intrusiveness and irritation to ads [11, 26]. People's responses to ads include affective, behavioral, and cognitive components [23, 94, 107]. Here, the affective component includes irritation and entertainment elicited by an ad [26], the behavioral component includes pre- and post- ad purchasing behavior [94], and the cognitive component includes factors like informativeness of an ad [23]. Relatedly, Bronner et al. studied the relationship between mood and ad effectiveness [13].

Parallelly, a body of research notes how time of day may affect the variety-seeking behavior of individuals [34]. In fact, circadian orientation and time of day are known to associate with an individual's depth of information processing with respect to ads [15]. Prior research studied how ad effectiveness varies with time of day by different age groups of individuals [32], and Tellis et al. studied the microeffects of time, content, and duration on ad effectiveness [98]. Relatedly, Kapoor et al. noted the promises of just-in-time recommendations in online platforms [45, 46]. Taken together, these studies explain how ad consumption is dependent on several contextual and psychological factors.

While the role of context in explaining ad effectiveness has been extensively studied in offline and traditional forms of media, it still remains an underexplored avenue in the space of social media and online platforms. In fact, with the emergence of newer forms of media and content delivery, it is important to assess contextual factors and accordingly improve content delivery to ensure better user experience [67]. Further, due to the lack of a comprehensive understanding of how users consume ads on these new online content delivery platforms coupled with ubiquitous technological affordances (such as smartphones and wearables), ad allocation on social media is still largely driven by only static rules and content-related personalization. Our work aims to address this gap in theory and practice by examining ad consumption with respect to time and momentary factors on Snapchat. We further simulate an experiment that evaluates the efficacy of our context-centric factors in making effective ad allocations.

### 2.3 Social Media Behaviors and Observational Data

A rich body of research reveals how social media activities reflect people's offline routines and behaviors [17, 36, 63]. Social media behaviors can potentially reveal naturalistic patterns of behavior, cognition, psychological states and social milieu, both in real-time and across longitudinal time [31, 58]. Prior work has also harnessed social media to infer individual-centric attributes ranging across personality traits and wellbeing using machine learning and computational linguistics [33, 79, 80, 92, 95]. Kosinski et al. used Facebook Likes to predict a range of sensitive personal attributes including sexual orientation, ethnicity, personality, intelligence, addictive behavior, age, and gender.

In the related problem space as ours, social media behaviors can explain ad consumption and vice versa [20, 96, 106, 108]. Kim et al. investigated the antecedents of clicking ads on Facebook [49] and Mao and Zhang studied the factors associated with users' intention to click on social media ads, particularly around content, media, and individual-related factors [61]. Prior work has also examined social media ads with respect to perceived informativeness, entertainment, and irritation [49, 62]. and Youn and Kim examined reactance related factors of avoiding ads on Facebook [108].

In general, the effect of certain changes or interventions is examined using causal-inference approaches. These approaches draw motivation from epidemiological research settings of randomized controlled trials (RCTs): participants are randomly assigned to a treatment and a control group where the former receives a drug, and the latter receives a placebo, and then changes are measured in the two groups to quantify the effect of the drug [37]. Similarly, understanding user behavior on a platform due to platform-based interventions are best studied with experimental and A/B test approaches [56, 70]. However, such approaches bear caveats. For instance, experimental studies that seek participant consent can be limited by concerns of observer effect [1] — participants may modulate their otherwise normal behavior with the awareness being monitored or observed. Alternatively, experimental research conducted without participants' awareness are deemed unethical especially in the human-centered research paradigm [43, 68]. For example, the Facebook emotion contagion study did not inform the participants that their feeds would be modified for research [53]. While this work was successful in uncovering valuable insights regarding people's affective behavior on social media, this work was heavily critiqued on ethical grounds [43]. Further, experimentation without apriori awareness of impact on participants may lead to long-term negative consequences for both platforms and individuals.

Consequently, in problem settings where experimental approaches may be infeasible or unethical, researchers have conducted observational studies. While observational studies cannot guarantee "true causality", they are designed in a way to minimize confounds and to investigate longitudinal data in providing stronger evidences than naive correlational analyses [41]. These studies can also benefit future randomized experiments where no preferred treatment is known apriori [88]. Recently, this kind of study has also generated interested in HCI, social, and behavioral science, including that using social media data [19, 47, 75, 89, 91, 110]. Notably, De Choudhury et al. examined the shifts in suicidal ideation tendencies in online communities [19] and Culotta estimated county health statistics using Twitter data [17, 19]. Of our particular interest is Saha et al.'s work which motivates us to operationalize social media behavioral measures such as activity, interactivity, and interaction diversity, whose relationship we examine with ad reception in our study [92].

Our work draws motivation from the success of observational data and quasi-experimental study design to understand user ad consumption on social media. Besides, we also note that our study values the importance of a contextualized person-centric design which is not only an improvement over one-for-all or generalized approaches but also stays clear of using demographic and trait-based information of users. While using such information may though improve machine

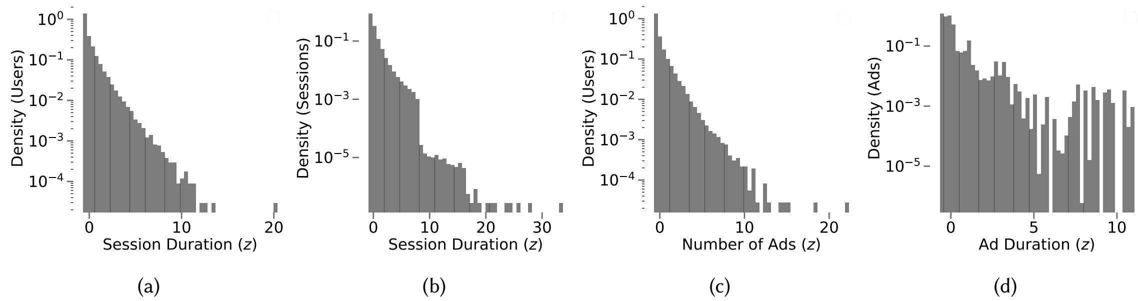



Fig. 1. Distribution of data by $z$-scores of (a) Total duration on platform per user, (b) Duration per session, (c) Number of ads per user, (d) Ad duration per ad.

prediction accuracy in some problem settings, these approaches could be exclusionary, discriminating, privacy-intrusive, and unethical [40, 81]. Rather, our study design incorporates dynamic platform behaviors to draw insights corresponding to user strata exhibiting similar combinations of platform behaviors.

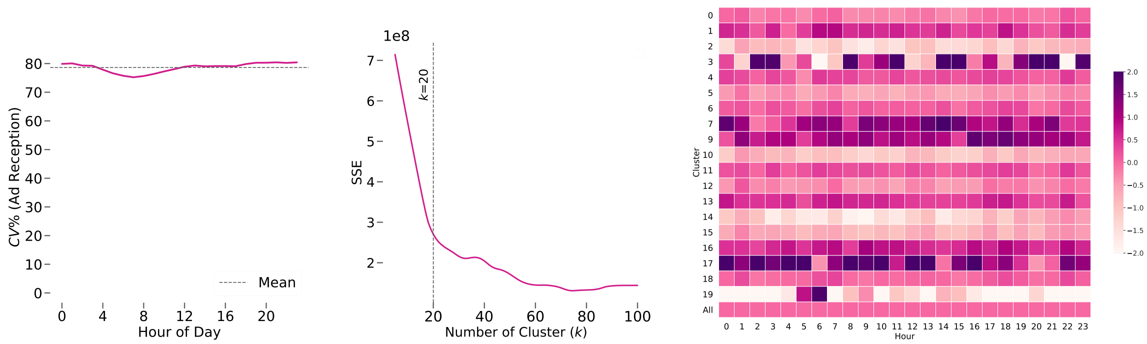
## 3 DATA

We conduct our study on the Snapchat platform. Snapchat is an online social and instant messaging platform that enables users to share and interact with others using ephemeral content, including text, images, videos, and other forms of multimedia. Snapchat is particularly very popular among the youth, with 73% of the 18-25 age demographic in the U.S. being Snapchat users [77]. Snapchat provides a Discover Feed where users can find and view recommended content in tiled story format from news publishers, brands, and content providers, such as ESPN, Wall Street Journal, Daily Mail, etc. Users can browse through these tiles, and when on a tile, they can consume, skip, or advance to the next recommended content. Snapchat's Discover Feed is design-wise similar to content-feed of Facebook, Twitter, or Instagram [2]. Discover Feed also shows ads which contribute to ∼98% of Snapchat's revenue [85]. As on most other platforms, users can watch an ad on Snapchat as long as they are interested, skip if they are uninterested or want to move on to other content, and/or swipe up (considered an ad click) if they are particularly interested to know more.

We scope our study to understanding ad consumption on Snapchat's Discover Feed. We first obtain a random sample of 100,000 users who have been active on Snapchat at least once every day for more than three months between December 17, 2019, and February 24, 2020. For these users, we obtain their longitudinal activity on the platform in the same period. Among these 100K users, this paper studies the data of 78,187 users' data who used the Discover Feed in this time period. We define each session as a continuous interval of time a user spends on the Snapchat app, or closes and opens it back within 15 seconds. Fig. 1 shows the $z$-scores of distributions of our dataset.

### 3.1 Preliminary Analysis

Adopting a definition of *ad reception* as the ratio of duration ads are watched over total duration of ads, we first conduct a preliminary analysis to understand how ads are consumed on the Snapchat's Discover Feed with varying hours in a day. For this, we measure the coefficient of variation (CV) of ad reception for each hour of the day. CV, expressed as a percentage, is the ratio of standard deviation to mean, quantifying the amount of variability with respect to the mean of the distribution — higher values indicate higher variability. We find that the CV per hour averages at a high 78.6% (stdev.=1.6), suggesting that users have high variance in ad reception by hour (ref: Fig. 2a).




(a) (b) (c)

Fig. 2. (a) Coefficient of Variation (CV) of ad reception across users in dataset by hour of day, (b) Elbow plot to determine optimal $k$ in $k$-means clustering: dotted line represents approximate location of elbow, (c) Ad reception per cluster of users by hour (values are $z$-transformed): darker colors indicate greater ad reception.

Further, to visually examine the above variability in ad reception by hour and by user, we cluster users on their aggregated activities on the platform (such as the number of app opens, frequency, and amount of posting and consuming content on the Snapchat platform). We adopt $k$-means clustering ($k$=20) where the number of clusters is roughly determined using the Elbow heuristic [93] (Fig. 2b). Figure 2c plots a heatmap of the mean $z$-score of ad reception by hour, with each cluster of users on the vertical axis. We find that the bottom-most row in the heatmap or the ad reception at an overall level barely shows any variance across hours. However, the same does not hold true if we look at the rest of the rows with shades of light and dark distributed throughout. This suggests that ad reception shows different patterns both between and within clusters across hours.

These preliminary analyses motivate us to investigate if users have different "preferred hours" of ad consumption (or times when ads would plausibly be less disruptive), and that clustering (or stratifying) users with on-platform attributes can provide key insights regarding ad consumption. These attributes do not use a user's demographic and trait-based information, and therefore, can be argued to be more privacy-preserving than traditional forms of user-profiling [76]. Concretely, any interventions using these insights would not need to access demographic or other personal data.

## 4 AIM 1: AD TIMINGS AND AD EFFECTIVENESS

### 4.1 Study Design and Rationale

Ads can disrupt a user's normal course of action on a social platform [4, 87]. Prior work has explored methodologies to improve ad effectiveness by showing personalized advertisements to individuals, where major approaches have largely focused on user interests and content-based personalizations (see Section 2). In this regard, context- and time- driven factors have remain largely unexplored, particularly on social media. Motivated by the role of context and time of day effects [15, 32, 98] and initial insights from our preliminary analysis, we hypothesize that different users have different preferred times of ad consumption on the platform.

Ideally, such a problem would be best examined in an experimental or A/B test setting; however, these methods have caveats [37]. For example, experimental allocations of ads may lead to unintended consequences such as changing platform experiences and risks of long-term perceptions about the platform. Again, these approaches are sensitive to particular parameters and thresholds, such as what quantity of ads can be shown and when — which remain unknown apriori to experimentation. Given these considerations, we draw on quasi-experimental approaches on observational data to understand the effect of ad allocations with respect to time preferences of users [88]. In particular, we adopt a causal framework based on matching, which simulates an experimental setting by controlling for as many covariates as possible [41]. This approach builds on the potential outcomes framework, examining if an outcome is caused by a treatment $T$ by comparing two potential outcomes: (1) $Y_i$ when exposed to $T$ ($T = 1$), and (2) $Y_i$ if there was no $T$ ($T = 0$). Because it is impossible to obtain both kinds of outcomes for the same user, this framework overcomes this challenge by estimating the missing counterfactual for a user based on the outcomes of a matched user — a user with similar distribution of covariates but differing treatment status. We employ stratified propensity score analysis [75, 91] to match users and examine ad outcomes in Treated and Control groups of individuals. This section describes the methodological considerations and approach in detail.

This paper communicates our insights using $z$-score-transformed quantities from raw data metrics due to privacy and sensitivity reasons. Importantly, $z$-scores are not sensitive to inconsistent magnitudes of absolute values, making normalized comparisons across multiple measures feasible. By definition, $z$-scores quantify the number of standard deviations by which the value of a raw score is above or below the mean. Similar standardization techniques have been adopted in prior social media studies [31]. $z$-scores are calculated as $(x - \mu)/\sigma$, where $x$ is the raw value, $\mu$ and $\sigma$ are respectively the mean and standard deviations of the population. Here, we obtain population $\mu$ and $\sigma$ on the entire data per measure. We interpret positive $z$-scores as values above the mean, and negative $z$-scores as those below the mean.

### 4.2 Defining Outcomes: Ad Effectiveness

A causal study typically measures the *effect-of-a-cause*, and the effect is measured in terms of changes in *outcomes*. Our work measures outcomes in terms of *ad effectiveness*. We draw motivation from prior research that ad effectiveness is quantified as a function of how interested people feel in watching an ad, and what actions they take following their consumption of the ad (such as buying the product, or other behavioral markers indicating their interest in the product) [71, 101]. On the basis of this, we operationalize ad effectiveness using two measures — 1) *Ad Reception* or the proportion of time ads are watched over the total duration of ads in a session and 2) *Ad Click Through Rate (CTR)* as the proportion of ads that were clicked on (or that users swiped up on, in the case of Snapchat).

### 4.3 Defining Baseline and Measurement Periods

We aim to measure ad effectiveness while conditioning on how ads were allocated to the users. For this purpose, we draw upon recent causal inference research on observational social media data [90], to define a Baseline and a Measurement period in the longitudinal timeline of each user (schematically represented in Fig. 3). In the Baseline period, we aim to compute how users consumed ads shown at different hours of day (and weekdays). This allows us to obtain the preferred hours of ad consumption for each user. Then, in the Measurement period we measure the effect of showing ads in preferred hours by minimizing the confounds due ad quantity and user engagement and behavior on the platform. We choose to split the longitudinal timeline of 78Kusers before and after January 10, 2020, leaving us with roughly three weeks of data in the Baseline period and six weeks of data in the Measurement period for each user.

### 4.4 Defining Treatment and Treated & Control Users

As we examine the effect of timing ad allocations, our study design adopts *Treatment Dosage* on the basis of preferred times of ad consumption. We operationalize treatment dosage on how similarly (or differently) ads were allocated in the Measurement period with respect to a user's high (or low) hourly ad consumption in the Baseline period. This builds on the notion that a user who consumed ads well at $H_1$ hours and poorly at $H_2$ hours during the Baseline period, would

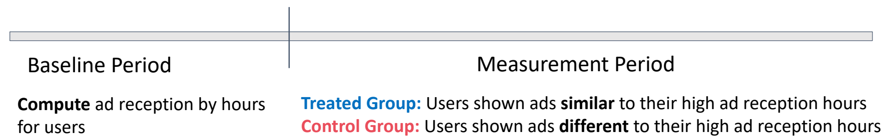



Fig. 3. A schematic representation of splitting user longitudinal timelines into baseline and measurement periods.

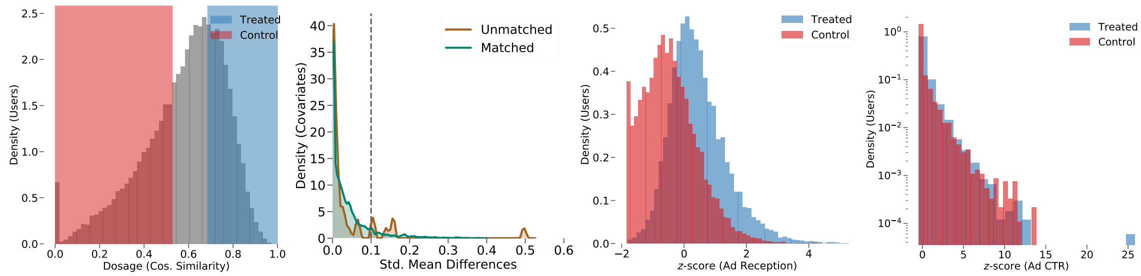
show similar consumption patterns when ads are allocated to them at $\mathsf{H}_1$ and $\mathsf{H}_2$ hours in the Measurement period. For each "hour of the day" in each "day of the week" (henceforth, referred to as hour-weekday pair), we compute an aggregated average of ad quantity (number of ads normalized by number of browsed content tiles) and ad reception per user separately in the Baseline and Measurement periods. First, we obtain the hour-weekday wise vector of ad reception for each user in the Baseline period ($v_1$). Then, we obtain the hour-weekday wise vector of ad allocation for each user in the Measurement period ($v_2$). Finally, we define treatment dosage as the cosine similarity of vectors $v_1$ and $v_2$, computed per user. Essentially, the greater the dosage, the greater is the likelihood of ad allocation (in Measurement period) during a user's preferred ad reception hours (as inferred from ad consumption behavior in the Baseline period).

For a better understanding of the effect of showing ads at preferred hours, we binarize the dosage into *treatment* and *no-treatment* based on various thresholds of percentile splits, creating "high similarity" (or Treated) and "low similarity" (or Control) groups. By varying the thresholds to binarize dosage into treatment and no-treatment, we find that our results are not sensitive to the choice of these splits of dosage. For clarity, we report and describe our results using the definition of treatment as the first tertile of dosage and no-treatment as the last tertile of dosage — which leads to 16,501 Treated and 16,501 Control users in our dataset (ref: Fig. 4a). Later, in Section 7, we revisit the robustness of our findings with respect to several combinations of dosage. The same section also examines dosage as a continuous variable and studies its relationship with ad outcomes.

### 4.5 Matching for Causal Inference

*4.5.1 Matching Covariates.* Matching aims to control for *covariates* so that the effects of treatment are examined between two comparable groups of users [41]. We note that ad outcomes can be confounded by factors such as how long someone stays in a session, or how do they engage on Snapchat, or even how much ads were allocated. To mitigate such confounds in our analyses, we adopt an approach called matching — when conditioned on high dimensional covariate data, matching can minimize biases compared to naive correlational analyses [41]. Our approach controls for a variety of covariates so that the compared Treated and Control show similar baseline behaviors. Drawing on prior work [47, 75, 92], we use 41 covariates. The first set of covariates are based on aggregated Baseline data spanning across the number and frequency of app opens, social interactions, different types of interactions, etc. We also include a second set of covariates obtained from the Measurement data based on the number of sessions, average duration of sessions, and the average number of ads exposed, to ensure that matched Treated and Control users had comparable engagement on the platform and were exposed to similar quantity of ads.

*4.5.2 Stratified Propensity Score Analysis.* We use matching to find pairs (generalizable to groups) of Treated and Control users whose covariates are statistically very similar to one another. The propensity score model matches users based on their *likelihood* of receiving the treatment, or the propensity scores. Stratified matching potentially overcomes the challenges of exact one-to-one pair matching which can lead to biases [50]. Our stratified matching approach




(a) Dosage (b) Matching Balance (c) Ad Reception (d) Ad CTR

Fig. 4. Distribution of: (a) Dosage: cosine similarity of ad reception in Baseline period and ad allocation in Measurement; (b): Balance of covariates in matching; (c&d): Ad outcomes (Ad Reception and Ad CTR): Treated users show greater ad outcomes on an average.

groups users with similar propensity scores into strata [47]. Every stratum, therefore, consist of users with comparable covariates. This approach allows us to isolate and estimate the effects of the treatment within each stratum.

To compute the propensity scores, we build a logistic regression model that predicts a user's binarized treatment status (0 for Control and 1 for Treated) based on their covariates. We segregate the remaining distribution into 100 strata of equal width — and discard those strata containing less than 50 users which further ensures that our causal analysis per stratum remains restricted to a sufficient number of similar users, and therefore is minimally biased [91]. This leads us to a final matched dataset of 92 strata consisting of 16,003 Treated and 15,682 Control users in total.

*4.5.3 Quality of Matching.* To test that our matching yields statistically comparable Treated and Control users, we evaluate the balance of covariates. For each covariate, we compute the standardized mean difference (SMD) in the Treated and Control groups in each of the 92 valid strata. SMD calculates the difference in the mean covariate values between the two groups as a fraction of the pooled standard deviation of the two groups. Two groups are considered to be balanced if all the covariates reveal SMD lower than 0.2 [47, 97]. This condition is satisfied by a majority of the covariates in our matched datasets, and we obtain a 18.9% reduction in the SMDs of matched from unmatched samples ($t = 0.19$, $p<0.05$) (Fig. 4b). Therefore, we can consider our matching to yield balanced Treated and Control groups of users that allow our ensuing analyses to be controlled on observed covariates.

### 4.6 Measuring Treatment Effect

To examine the effect of timing ads based on users' Baseline-inferred preferred ad times (or treatment), we compute the differences in the outcomes (ad effectiveness) between the matched Treated and Control users in the Measurement period. We compute these differences in terms of effect size (Cohen's $d$) and paired $t$-tests which also helps us to evaluate the statistical significance in differences. We also conduct Kolmogorov-Smirnov (KS) test, which tests against the null hypothesis that the outcomes in the Treated and Control groups are drawn from the same underlying distribution.

To quantify the *effect* of treatment, we measure the Relative Treatment Effect (RTE) per outcome measure in every strata, as the ratio of the likelihood of the outcome in the Treated group to that in the Control group [47, 91]. Next, using a weighted average across all the strata, we obtain the mean RTE of the treatment per outcome measure. An outcome RTE greater than 1 would mean that the outcome is greater in the Treated users in the Measurement period — or in our case, that allocating ads according to preferred timing increases the likelihood of ad effectiveness.

Table 1. Summary of mean $z$-scores in Treated and Control groups along with Relative Treatment Effect (RTE), Effect Size ($d$), paired $t$-test and $KS$-test. Statistical significance reported after Bonferroni correction (* $p$<0.05, ** $p$<0.01, *** $p$<0.001).

| Outcome | Tr. (z) | Ct. (z) | RTE | d | t-test | KS-test |
|---|---|---|---|---|---|---|
| Ad Reception | 0.29 | -0.33 | 1.49 | 2.27 | 15.28*** | 0.83*** |
| Ad CTR | 0.06 | -0.09 | 1.51 | 1.11 | 7.52*** | 0.55*** |

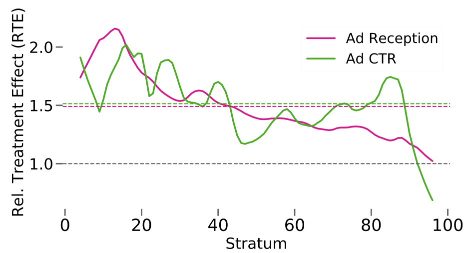



Fig. 5. Ad outcomes per matched strata of users.

### 4.7 Findings: What is the Effect of Timing Ads?

Fig. 4c and 4d show average changes of ad outcomes in Treated and Control users — indicating higher ad outcomes for Treated users. Table 1 reports the differences in outcomes of the Treated and Control users in terms of RTE, Cohen's $d$, paired $t$-test, and KS test. These tests indicate statistically significant outcome differences in Treated and Control users. The high magnitude Cohen's $d$ values ($d$>0.5) and $t$-statistic, along with the positive signs indicate that the Treated group shows significantly higher outcomes than the Control group. Additionally, the fact that ad reception shows an RTE of 1.49 and ad CTR shows an RTE of 1.51, indicates that timing ads (treatment) leads to more effective ad outcomes (RTE>1). Fig. 5 reveals that the RTE>1 holds for a significant majority of user strata for both ad outcomes.

This reveals that even though matched Treated and Control are very similar to each other on baseline behavior and ad consumption, along with Measurement period's ad exposure and platform activities (as per matching), they show very different receptivity to ads in the Measurement period. Therefore, we draw two inferences. First, the short baseline period of understanding on-platform activities of users can lead us to passively infer *preferred times of ad consumption*. Second, when users are shown ads at preferred times, they are more likely to be receptive to the ads — or they would watch the ads longer, and are more likely to click on them. On the other hand, ads shown during less preferred hours may not only be worse received but may also elicit frustration or seem intrusive to users [26].

## 5 AIM 2: USER BEHAVIOR AND AD EFFECTIVENESS

A core finding of Aim 1 is that users are more likely to be receptive to ads if ads are shown according to their preferred times. However, allocating ads in this way may not always be feasible — for example, a user may show inactivity or hyperactivity during their preferred times, each of which may lead to extremes of no ad or too many ads to be allocated, thereby affecting either or both of platform revenue and user experience. Our Aim 2, in particular, examines alternative means of effective ad allocation which is robust to a user's (unknown apriori) future platform activity. That is, we investigate how in-the-moment user behaviors associate with a user's ad consumption at varying times of a day, which would provide insights to conduct context-centric ad reallocations considering factors beyond time of the day.

First, for every user, we define each hour as a *less-preferred* or a *more-preferred* hour of ad effectiveness on the basis of whether the hour falls on either side of the median ad reception of the user in their Baseline period. Consequently, we label every session in the Measurement period of a user to be either during *less-preferred* or *more-preferred* hours of ad exposure for the same user. We operationalize multiple in-the-moment user behaviors on the platform and examine the relationship of these measures with ad effectiveness, conditioned on preferred times of ad exposure. This section explains the relationship of ad effectiveness with on-platform user behavior.

### 5.1 Examining the Relationship with Ad Effectiveness

We aim to understand how can ad allocations be improved for a particular session beyond the timing of the ad. Therefore, given a user and a session, we conduct a two-fold examination of ad effectiveness based on 1) if the session is during a more-preferred ad time of the user, and 2) characteristics of the session in terms of on-platform user behaviors. For sessions in less-preferred times, we conduct a paired $t$-test of in-the-moment session characteristics of those sessions when ad effectiveness was higher than median for a user (or high effective) and those sessions when the ad effectiveness was lower than median for the user (or low effective). Similarly, for more-preferred times, we conduct a paired $t$-test of session characteristics of high ad effective and low ad effective sessions per user. Essentially, a statistical significance in these comparisons would indicate that in-the-moment user behaviors associate with ad effectiveness within and outside preferred times of users. The sign of $t$-statistic would indicate the directionality of the measure with ad reception, positive values indicate a positive association and negative values indicate the opposite.

### 5.2 Ad Effectiveness and Passively Inferred In-the-moment Behavior

Ad consumption is known to be a function of people's momentary psychological and behavioral states (Section 2). For instance, ad consumption (and more generally, content consumption) is a function of what an individual does, or how active or tired they are at a particular block of time [13, 98]. As a proxy of behavioral and psychological states, we draw on prior work to operationalize a variety of passively inferred in-the-moment states on the Snapchat platform [92]. We explain these below. We first motivate and operationalize each of these in-the-moment behaviors and follow that with our observations with respect to the relationship with ad effectiveness in each. We are particularly looking for insights to recommend ad allocations with respect to preferred times with minimal compromise on ad effectiveness, i.e., 1) when to increase ads during less-preferred times and 2) when to decrease ads during more-preferred times. Table 2 summarizes the differences in high and low ad effective sessions by preferred times of users.

**Duration.** We operationalize duration as the length of time a user spends in a session. We note that although longer sessions indeed allow the platform to show more ads, longer sessions also plausibly correlate with a user's leisure times, or when they are less involved with offline activities. Table 2 shows that $t$-test on high and low effective ads during less-preferred ($t$=7.27) and more-preferred ($t$=15.05) times is positive with statistical significance, suggesting that ad effectiveness positively associates with length of session. Therefore, a recommendation at less-preferred times would be to increase ads in longer sessions, and that at more-preferred times would be to decrease ads in shorter sessions.

**Activity.** We operationalize activity as the frequency of touch actions (excluding text-typings) in a session. A more active user may be more likely to skip ads and move on to a different content. For activity, we find that $t$-test for both less-preferred times ($t$=−5.26) and more-preferred times ($t$=−3.78) is negative with statistical significance. This indicates that activity shares a negative relationship with ad effectiveness. Therefore, a recommendation for less-preferred times would be to increase ads during low-activity sessions, whereas a recommendation for more-preferred times would be to decrease ads during high-activity sessions.

**Interactivity.** One way to study user behavior on an online social platform is measuring a user's degree of interactiveness in terms of posting, responding, and consuming content [92]. We operationalize interactivity as the ratio of content created to content consumed within a session. Here, content creation on the Snapchat platform includes creating stories, posting updates, and sending and replying to chat messages, while content consumption includes viewing others' stories and updates, and browsing through different content within a session. Table 2 shows that for both less-preferred times ($t$=−3.18) and more-preferred times ($t$=−5.25), $t$-statistic is negative. This suggests that interactivity negatively associates with ad effectiveness, or higher the interactivity of a user in a session, lower is their ad reception. A plausible interpretation of our Interactivity results is that, when users encounter ads on Discover Feed during highly interactive sessions, they might feel particularly disrupted and have specific actions available (e.g. chatting with a friend), motivating them to skip ads. Our findings, therefore, recommend, to increase ads during low-interactivity sessions of less-preferred times, and decrease ads during high-interactivity sessions of more-preferred times.

**Interaction Diversity.** Another form of social media behavior corresponds to the diversity of interactions an individual conducts during a session [92]. We operationalize interaction diversity as the standard deviation of time spent in each kind of activity conducted in a session, where in-session activities range across sending or replying to chats, viewing and posting updates, etc [36]. Like the above, both less-preferred ($t$=−5.61) and more-preferred ($t$=−4.78) times, exhibit statistically significant negative $t$-statistics. This suggests that interaction diversity negatively associates with ad effectiveness. Our findings, therefore, recommend, to increase ads during low interaction diversity sessions of less-preferred times, and decrease ads during high interaction diversity sessions of more-preferred times.

**Distractedness.** Although there is no accurate means to passively infer how distracted a user is, we hypothesize that a distracted user (with respect to the app) would plausibly conduct more non-app related activities during a block of time, such as switching to another app, or attending a phone call, or doing something else offline and returning back to the app, etc. We operationalize distractedness as the quantity of application opens and closes during a session (recall that a session does not end until a user has been inactive for 15 seconds). For less-preferred times, we find no statistical significance in the differences of distractedness in high and low ad receptive sessions. However, at more-preferred times, we find a positive $t$=2.01 with statistical significance. This might mean that when users visit the platform being less distracted, they are plausibly doing something with a "particular purpose" and may not be willing to consume ads despite being at their preferred times of ad consumption. This suggests a recommendation that during more-preferred times, ads can be decreased in less-distracted sessions.

**Extra-socialness.** Platforms such as Snapchat, Instagram, and Facebook also provide features that are not necessarily "social" or "interactive", e.g., playing games, using a camera and applying filters or lenses on their photos, etc [36]. We operationalize the ratio of time spent on these activities to the total session duration as extra-socialness of a session. Comparing extra-socialness of different ad reception sessions at less-preferred times, we find no statistical significance, and at more-preferred times, we find $t$=−1.35 with significance. This indicates that, at more-preferred times, extra-socialness negatively correlates with ad effectiveness, or when a user is interested in non-social platform activities, they are less likely to be receptive to ads, finding them disruptive. Correspondingly, a recommendation would be to decrease ads during high extra-social sessions at more-preferred times of users.

**Summary of Insights and Recommendations.** The above observations suggest that ad effectiveness positively correlates with duration and negatively correlates with activity, interactivity, and interaction diversity for any time; positively correlates with distractedness and negatively correlates with extra-socialness at more-preferred times. Therefore, recommendations for *increasing ad allocation at less-preferred times* are sessions with 1) high duration, 2) low activity, 3) low interactivity, and 4) low interaction diversity. On the other hand, recommendations for *decreasing ad allocation at more-preferred times* are sessions with 1) low duration, 2) high activity, 3) high interactivity, 4) high interaction diversity, 5) low distractedness, and 6) high extra-socialness. While we study the relationship between on-platform momentary behaviors and ad effectiveness here, approaches to infer momentary behaviors in real-time or

Table 2. Summary of in-the-moment user behaviors with respect to ad effectiveness on the same user. Statistical significance is conducted using paired $t$-tests and $p$-values are reported after Bonferroni correction (* $p$<0.05, ** $p$<0.01, *** $p$<0.001). For significant rows, **violet** bars represent **positive** magnitudes, whereas **orange** bars represent **negative** magnitudes.

| Measure ↓ | Less preferred Time | | | | More preferred Time | | | |
|---|---|---|---|---|---|---|---|---|
| Effectiveness → | High | Low | t-test | p | High | Low | t-test | p |
| Duration | -.031 | -.047 | 7.27 | *** | -.004 | -.038 | 15.05 | *** |
| Activity | -.006 | .012 | -5.26 | *** | -.010 | .003 | -3.78 | ** |
| Interactivity | .001 | .014 | -3.18 | ** | -.007 | .012 | -5.25 | *** |
| Interaction Diversity | -.012 | .007 | -5.61 | *** | -.018 | -.003 | -4.78 | *** |
| Distractedness | .010 | .011 | -0.22 | | .017 | .009 | 2.01 | * |
| Extra-socialness | .004 | .005 | -0.21 | | .005 | .010 | -1.35 | * |

apriori are beyond the scope of this study. However, these can be implemented using real-time dynamic rules (e.g., if the current session duration or session interactivity is already higher than the user's median at a given time) or using predictive machine learning techniques [6, 51].

## 6 AIM 3: EVALUATING THE EFFICACY OF INSIGHTS

Now that we derived insights in Aim 1 and 2 regarding ad consumption based on ad allocations at preferred hours and on-platform behaviors respectively, we ask how these insights would influence business value? We conduct a simulation experiment of increasing and decreasing ad allocations based on recommendations guided by our first two research aims. Additionally, we were concerned that intervening in the status quo of the ad allocation process could create concentrated ad allocations, i.e. certain users bearing the burden of high "ad load". Thus, this additional investigation aims to evaluate how simulated interventions might affect the fairness of ad allocations in terms of the concentration of ad load among users. This simulation experiment can inform an ad allocation system about weighing in ad allocations to users who are at extremes of ad exposure to balance the quantity of overall ad exposure across all users on the platform, i.e., a more balanced but effective allocation of ads.

### 6.1 Simulating a Balanced and Effective Ad Reallocations

We first quantify the *normalized ad quantity* per user as the ratio of the total number of ads over the total number of contents (or Discover Feed story) seen by the user. Fig. 6a shows the min-max scaled distribution of normalized ad quantity within our Measurement dataset. We identify *high and low ad-exposed users* as the top and bottom quartile of normalized ad quantity. Because these users are at extremes of ad exposure during a particular period, a platform would ideally like to first change the ad quantity of these users to similarly balance out normalized ad quantity across all users. Accordingly, we *simulate* new ad distributions by decreasing the ad quantity of high ad-exposed users and increasing the ad quantity of low ad-exposed users in the following three ways:

**Preferred Time based Reallocation.** In this simulation approach, we use recommendation solely from Aim 1, i.e., for high-ad exposed users, we decrease the number of ads in less-preferred hour sessions (in hours where ad reception in Baseline period is lower than the median) by 90% per session and allocate the difference in quantity of ads proportionately across the preferred hour sessions (ad reception in Baseline higher than the median) of the low ad-exposed users.

**Session Activity-based Reallocation** In this simulation approach, we use the recommendations from Aim 2 to modulate the ad load of high and low ad exposed users. In high ad-exposed users, we decrease the ad quantity in a union of sessions with low duration (lower than bottom 25 percentile), high activity, high interactivity, high interaction diversity, etc. (higher than top 25 percentile) by 90% per session. Similar to the above, we allocate the difference in quantity of ads proportionately in sessions of low ad-exposed users, in those sessions with low activity, low interactivity, low interaction diversity, etc. (lower than bottom 25 percentile).

**Baseline Reallocation.** In the third simulation approach, we build a baseline reallocation which does not use the insights from the previous two research aims. This is aimed to sort of emulate a status-quo of platforms that follow fair and balanced ad allocations — when users are identified with extremes of ad exposure in real-time, their ad exposure is modulated to balance in the upcoming period. In the baseline reallocation, we randomly select $n$ sessions from high ad-exposed users and decrease their ad load by 90% per session, and allocate the difference in randomly selected sessions of low ad-exposed users. We choose $n$ as the same number of ads reallocated in the above two allocations, as the baseline reallocation is to compare against the two other reallocation strategies. To eliminate any effect due to chance, we build 1,000 permutations of different $n$ sessions where ads are manipulated.

### 6.2 Evaluating Ad Reallocations

We evaluate ad reallocations on the basis of *ad value*, which is a function of how effective ads are in a session. We measure *ad value* as a product of ad reception and the number of ads in a session. Theoretically, ad value would be correlated with actual monetary value generated based on ad effectiveness [71, 101]. We note that our simulations are only within the limits of the observational data, and our measure of ad value assumes a user's ad reception in a session to be the observed value. However, it is likely that the counterfactual ad reception might change if the ads are actually reallocated — which remains unknown unless an actual experiment of ad reallocation is conducted.

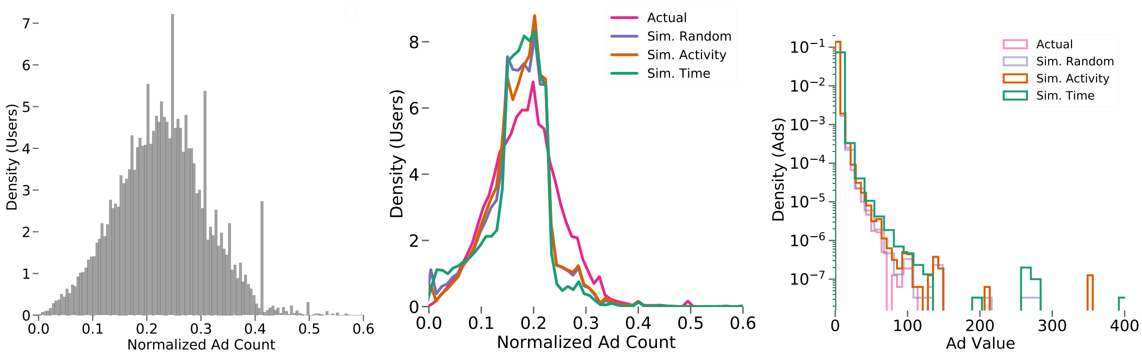
Fig. 6b shows the distribution of ads across users in multiple simulation strategies. We find that compared to the actual distribution, 1) the baseline simulation decreases the standard deviation by 9%, 2) the simulation by activity decreases by 6.5%, and 3) the simulation by preferred time decreases by 7.9%. Lower standard deviation suggests that all our simulation forms of ad reallocations result in a balanced allocation of ads across users.

Next, Fig. 6c shows the distribution of ad value by simulation strategies. We find that compared to overall ad value in the actual distribution: 1) the baseline simulation only marginally increases ad value by 0.07%, whereas 2) the simulation by activity strategy increases ad value by 2.78%, and 3) the simulation by time strategy increases ad value by 7.09%. Both of these percentage changes actually correspond to a significant increase in overall value considering the scale of userbase and volume of data and ads on the platform, e.g., ∼250M daily active users on Snapchat [109].

Finally, drawing on permutation test approaches [5], we iterate over the 1,000 permutations of baseline reallocations, to find the probabilities ($p$-values) of the ad value improvement in the baseline reallocation over the two insight-driven reallocations. We find these probabilities to be zero, suggesting that we can reject the null hypothesis that insight-driven simulations only beat the baseline ad reallocation by chance.

## 7 ROBUSTNESS OF FINDINGS

This section examines the robustness of our findings with respect to the researcher decisions we made in our study. First, we conduct methodological robustness tests on parametric choice and approach in our study design. Then, we theoretically contextualize our definition of "ad effectiveness" measures with respect to how it is defined traditionally in the literature — a success would provide criterion and construct validity to our study. Together, a convergence in findings along with theoretical grounding would ensure robustness and validity of our findings [59].




(a) Ad Exposure (b) Ad Exposure (c) Ad Value

Fig. 6. Distribution of a) normalized ad count by user in the actual dataset, b) normalized ad count as per simulations of ad reallocations: the density plot for each simulate reallocation is thinner in width compared to the actual distribution suggesting a more balanced ad allocation across users, c) ad value as per simulations: overall ad value is highest for simulated by time reallocation (Sim. Time) followed by simulated by behavioral activty based reallocation (Sim. Activity).

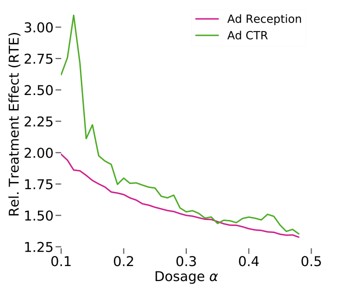



Fig. 7. RTE with varying dosage $\alpha$. For each $\alpha$, treatment is top $\alpha * 100$ percentile of dosage and no-treatment is bottom $\alpha * 100$ percentile of dosage.

Table 3. Coefficients of linear regression of relevant covariates as independent variables and ad reception as dependent variable, * $p$<0.05, ** $p$<0.01, *** $p$<0.001. For significant rows, **violet** bars represent **positive** magnitudes, whereas **orange** bars represent **negative** magnitudes.

| **Measure** | **Coeff.** | **p** | **Measure** | **Coeff.** | **p** |
|---|---|---|---|---|---|
| Treatment Dosage | 0.22 | *** | Num. App Opens | -0.06 | ** |
| Duration | 0.05 | ** | Interactivity | -0.11 | * |
| Activity | -0.04 | * | Interaction Diversity | -0.07 | * |
| Consumption | -0.03 | – | Distractedness | 0.01 | – |
| Curation | -0.03 | – | Extra-Socialness | 0.03 | * |

### 7.1 Methodological Robustness

**Binarizing Treatment Dosage Thresholds** Recall that our study design relies on chosen threshold of binarizing treatment and no-treatment (Section 4). We test if our findings hold robust for any other thresholds of treatment. For this, we vary the threshold of treatment dosage and re-conduct the entire analyses on measuring treatment effects (Aim 1), including matching and computing differences in outcomes for matched Treated and Control users. We vary the threshold parameter $\alpha$ in such a way that dosage (cosine similarity between Baseline ad reception and Measurement ad allocatino) in the top $\alpha * 100$ percentile is considered treatment, and that in the bottom $\alpha * 100$ percentile is considered no-treatment. Figure 7 plots the change in RTE of ad effectiveness measures, with respect to changing the threshold dosage $\alpha$. We find that the RTE of both ad effectiveness measures remain greater than 1 (along with statistical significance as per $t$-test and effective size), indicating that ads were more effective on Treated users or users who were shown ads at their preferred hours. We also find a roughly monotonic decrease in RTE with respect to increasing $\alpha$, suggesting the greater the similarity of ad allocations with people's preferred hours, the greater is the likelihood of ad effectiveness.

**Using Treatment Dosage as a Continuous Variable.** Another component of our work includes the decision to separately estimate the outcomes by binarizing the treatment dosage not only for better interpretability purposes but also to emulate conventional RCT or experimental approaches where typically one group is treated (e.g., drug) and the other group is not (e.g., placebo). While less likely, binarizing treatment might however introduce new biases in the analyses and may lead to misleading findings (e.g. a drug in low dosage may not be as effective as it is in high dosage, [39]). Therefore, we also test the findings if we consider the treatment as a continuous variable.

For this, we build a linear regression model that uses all the 42 covariates in our dataset, along with the treatment as independent variables and the ad reception as the dependent variable[1]. While this approach is not particularly "causal", it allows us to infer the relationship of the treatment and the outcomes. We eliminate correlated features using variance inflation factor (VIF) (threshold=10) [18, 69]. The regression model shows an adjusted $R^2$ of 0.87, and Table 3 reports the standardized coefficients of relevant variables. In particular, we find that the treatment dosage shows the greatest magnitude with a positive coefficient of 0.22 ($p$<0.05). This aligns with our matched and binarized treatment analysis that *treatment* (or showing ads during preferred hours) leads to a greater likelihood of ad effectiveness.

The consistency of results via different approaches reveals that our examination is not sensitive to the choice of treatment dosage parameter or our study design, but rather a reflection of ad consumption behavior on Snapchat.

### 7.2 Contextualizing within the Literature

As a final robustness check, we compare our results with that found in previous literature. Traditionally, ad effectiveness is defined as whether an individual buys a commodity following exposure to an ad [13]. If our observations of ad effectiveness match prior literature, we view that as criterion validity to our measures of ad effectiveness and construct validity to our study [74]. We test ad effectiveness in two ways: in terms of the time of day and as day of the week.

*7.2.1 Time of the day.* We construct our first hypothesis based on prior work comparing ad effectiveness at days and nights [98] that: *ad effectiveness is higher during the daytime compared to evenings or nights*. In our work, we first bucket a user's local time into day (6 AM–6 PM) and night (6 PM–6 AM). While we also attempted to build more granular buckets or even look at ad effectiveness over more continuous forms of time, we find effects are generally washed out given the across-user variability in ad reception, also evident in our preliminary analysis referring to Figure 2c, particularly in the bottom-most/"All" user row. Instead, binary buckets (day and night) provide us the opportunity to compare and contrast users' receptivity to ads between broader timespans of a day.

For both groups of Treated and Control users, we conduct paired $t$-tests between a user's ad outcomes during the day and during the night. Table 4 reports the differences in ad reception by the time of the day. First, understandably, any ad effectiveness for Treated users is higher than Control users as also reflected in our other analyses (Section 4). Next, we find that for both Treated and Control users, ad effectiveness is higher during the day than night with statistical significance and large $t$-statistics, replicating prior research on ad effectiveness [98].

*7.2.2 Day of the week.* Prior work compared ad effectiveness on weekends and weekdays [13], and saw differences, particularly on the basis that weekends are associated with more home, leisure, and family events that might elicit more pleasant effects in people's mood and correspondingly in their receptivity to ads. Therefore, we construct our hypothesis that: *ad effectiveness is higher on the weekends compared to weekedays*.

We statistically compare ad effectiveness on weekends and weekdays, as reported in Table 4. First, ad effectiveness is higher for Treated than Control users. Next, within Control users, we find that both forms of ad outcome are higher during the weekends than on weekdays. In contrast, in the case of Treated users, ad reception during the weekdays and weekends do not differ with statistical significance. This indicates support for our hypothesis in the case of Control users, and lack of support in the case of Treated users. These findings suggest that users in the Treated group already

[1]We also conduct linear regression with ad CTR as dependent variable, which leads to similar signs of coefficients

Table 4. Summary ad effectiveness ($z$-scores) by hour and weekday on the same user. Statistical significance is conducted using paired $t$-tests and $p$-values are reported after Bonferroni correction ($^{*}$ $p$<0.05, $^{**}$ $p$<0.01, $^{***}$ $p$<0.001). For significant rows, **violet** bars represent **positive** magnitudes, whereas **orange** bars represent **negative** magnitudes.

| Measure | Treated Users | | | | Control Users | | | |
|---|---|---|---|---|---|---|---|---|
| | Weekdays | Weekends | t-test | p | Weekdays | Weekends | t-test | p |
| Ad Reception | 0.17 | 0.12 | 0.95 | | -0.34 | 0.04 | -8.27 | *** |
| Ad CTR | -0.38 | 0.55 | -8.36 | *** | -0.64 | 0.47 | -9.32 | *** |
| | Day | Night | t-test | | Day | Night | t-test | |
| Ad Reception | 0.48 | -0.16 | 7.42 | *** | 0.08 | -0.40 | 7.22 | *** |
| Ad CTR | 0.43 | -0.20 | 4.18 | *** | 0.33 | -0.57 | 6.78 | *** |

see ads at their preferred time, and as a result the weekend/weekday effect is diminished. Therefore, whether timing ads likely plays a stronger role than day of the week effect — future research may be able to shine more light on this.

## 8 DISCUSSION

### 8.1 Theoretical Implications

Our work opens up discussions on understanding ad allocation and consumption in new forms of media. Traditional media (television, newspapers) drive ads dedicated to audience groups, and ad effectiveness typically measures the amount of product purchases (by ad influence). In recent times, not only with the increasing use of online social platforms, but also with the ubiquity of smart and personal devices, ads are allocated in various novel ways. The affordances not only enable platforms to customize and allocate ads in a personalized fashion, but also provide users with choices to skip and ignore ads. Importantly, negative perception towards ads can cause user fatigue and exacerbate their perception of the platform [12]. This calls for a need to better understand ad consumption and accordingly design robust and dynamic ad allocation strategies. Our work reveals that when ads are allocated in a better fashion by accounting for user- and context-centric factors (e.g., time and platform engagement), users could be more receptive to ads, which corresponds to prior observations regarding intrusiveness and likeability of ads [49, 62].

We have couched our observations in theories from marketing science, psychology, and cognitive sciences. Our work augments these bodies of prior research, that have largely studied how the content of ads matters in changing user experience on online platforms [4, 16]. Our work provides valuable insights regarding the importance of context and momentary factors in understanding ad effectiveness. In particular, our observations suggest that timing ads is a factor that cannot be ignored when allocating ads on social media.

Along similar lines, our work also reveals how blanket approaches of ad allocation or approaches based on average user behavior may not be as effective. For instance, these approaches typically assume a linear relationship between a user's time spent in a session and the number of ads shown: ads are shown after a fixed duration or after showing a fixed number of pieces of content. However, these approaches ignore the cognitive state of users which can vary due to time of day effects or due to users' daily routines, or even momentary psychological states such as feeling social or excited at a particular moment [32, 98]. Instead, our work finds that when ads are allocated by accounting for these factors, ad effectiveness is higher without compromising the user activities on the platform.

### 8.2 Practical and Design Implications

**Individual-centric Implications.** Our work has implications for making responsible and user-facing advertising — advertising that aims to not only increase platform revenue, but also minimizes user dissatisfaction caused by ads and therefore keeps the users better engaged with the platform [99]. By contributing towards the niche aspect of "preferred timings", our work provides an approach to balance ad effectiveness with user experience. Our approach can be used to minimize the privacy intrusions generally associated with targeted advertising: We can reduce the use of profiling to target ads, and obtain preferred ad timing on de-identified features and short-term data.

Because users typically dislike ads and do not like to share their data with platforms for ad targeting [72], they often use tools such as ad-blocking and private browsing that do not share cookies and browser history with advertisers. Some platforms disallow these privacy-preserving practices, forcing the user to trade off their data in order to access the content. Potentially, users may be more comfortable sharing only their momentary (session-level) data, and our work shows that platforms can make effective ad allocations by only using these minimal, momentary user data.

Further, the efficacy of recommending content on the basis of time and context we have shown, suggests design implications that take user agency and user preferences into account [99]. Recently, the HCI community has demonstrated the value of user-contributed preferences of notifications and interruptions [65, 78, 99]. In line with this, platforms could ask users which times of the day they are more inclined to view ads, and allocate ads accordingly. This approach will require more insights in potential ways users could game the system (users might provide preferred times of ads when they would likely not visit the platform).

**Platform-centric Implications.** Our methodology allows platforms to allocate ads in an effective, fair, and less-intrusive way. A recent survey revealed three major categories of ad dissatisfaction are intrusiveness, annoyance, and disruptiveness of ads [104]. Users often use ad blockers and other tools that prevent ads on online platforms [29, 70] — these approaches raise nuanced questions surrounding the sustainability of platforms surviving on ad-driven business models [4]. Consequently, to protect user base and minimize ad-based interruptions, some platforms are moving away from ad-based models to some form of subscription-based models [30, 105]. However, such models have their own caveats, such as inequity of information access on the internet, and online services could become a function of an individual's ability to pay. Our work suggests somewhat of a middle-ground: by optimizing ad timings and allocations when users are less likely to feel interrupted, platforms can consistently provide equitable content access and experience to users, and better sustain the ad revenue ecosystem, with less user dissatisfaction.

**Towards Fewer Ads.** Our study also has implications towards optimizing other forms of ad allocations on social media, including ad spacing and ad loading. One can draw an insight that if we can allocate ads optimally in an effective fashion, we can plausibly reduce the overall quantity of ads if certain revenue goals are already achieved with smaller quantity of, but better-allocated (timed) ads. In fact, this can help minimize practices such as non-skippable ads (ads which cannot be skipped) or forced ads (ads which prevent any content consumption without being watched). Solutions of minimal ad allocations would potentially be well-appreciated by the users, improving the general user experience, and potentially leading to higher user retention on the platform [64]. A better ad allocation strategy provides a method for platforms to judiciously serve effective ads. As a result, such platforms can become more attractive to users.

**Small Data and Privacy-sensitive Approaches.** Research highlights several biases in ad delivery [3, 55, 84]. For instance, demography and inferred-user interest-based targeting can be deemed privacy-intrusive, unethical, and surveillance-promoting [40, 81, 87]. In contrast, our work shows a novel means to increase ad effectiveness that does not lean on these critiqued approaches. Our approach only requires short-term user data (e.g. a few weeks) instead of using long-term historical data. Long-term data do not only increase privacy concerns, but are also less robust to changes in both human behavior and platform affordances [9]. Therefore, small-data-driven approaches that do not compromise on the user experience can open up new opportunities in ad and content recommendation.

**Other Content Recommendations.** While we primarily focus on ads, our work also has implications for other types of content. Our work provides general insights for when to show content to users. This could inform design strategies for recommendations and notifications for preferred user content. Prior work showed the value of context-aware recommendations for improving user engagement on mobile platforms [46, 65]. The on-platform behaviors we studied (particularly in Section 5) can guide designing such recommendations that take a user's momentary state into account.

**Implications for Experimental Approaches.** As we noted earlier, causal effects are best studied with experimental approaches, which however, come with risks, e.g., certain treatments (design changes) may affect the perception of users and impact user retention. Moreover, in the case of continuous treatments, it is often difficult to determine the appropriate dosage to experiment on. Our study adopts a quasi-experimental design to show that a particular treatment (timing ads) can be effective. Our computational framework also quantifies "preferred timings" based on observed ad outcomes in a small time period. Therefore, our findings can help to formulate appropriate parameters (or dosage cutoffs as shown in Section 7) to conduct careful experimental studies to verify and adopt design changes on platforms [41, 70].

**Substitute or a Complement on the Existing Ad Allocation System?** Lastly, we raise a critical point. We conduct our study on a system already optimized (in some form) for ad allocations. Therefore, our effects may be even bigger if we had a different baseline. One might argue that our study only builds on the top of the existing system which could already be privacy-intrusive and be using inferred user interests based targeting — the very points on which we discuss several implications above. In this regard, it indeed remains unexplored whether momentary and context-driven features will be adopted in practice to improve ad effectiveness. There is a risk that companies will stack profile and momentary state approaches (instead of replacing the profile-based approaches), which could potentially be more privacy invasive. However, our study encourages to consider and evaluate these alternative strategies that conduct responsible and non-invasive ad allocations. It would be immensely insightful if future research suggests that we can (or cannot) significantly minimize or even eliminate any sort of profiling and demographic based targeting approaches. Ad targeting is coming under increased scrutiny, and as companies and governments are putting in place more privacy restrictions [103], our approach can help future-proof the ecosystem of ad-based platforms. Overall, an important takeaway of this work illustrates the importance and feasibility of building complementary methodologies which simultaneously consider a user's privacy and optimize for user experience and business value for long-term sustainability.

### 8.3 Ethical and Privacy Implications

We note that our work bears ethical implications. Our work is predominantly motivated by the idea that we might move away from traditional forms of user profiling and the using static demographic and trait-based information that content recommendation algorithms infer and use, which can be biased and unfair [40, 81]. However, this work can be used to conduct new forms of user profiling on people's online behavior. Online platforms could (mis)use our approach to conduct newer and plausibly unknown forms of biased and intrusive ad targeting, e.g., if these algorithms incorporate not only "who someone is", but also "what do they do when". As Pandit and Lewis describe, "the use of personal data is a double-edged sword that on one side provides benefits through personalisation and user profiling, while the other raises several ethical and moral implications that impede technological progress" [76]. Therefore, we need to think about balancing the costs and benefits of these approaches, along with implementing the systems in ethical and privacy-preserving fashion. While arguably anonymized and on-platform in-the-moment behavior is more ethical and less biased compared to demographic, static, and prior-assumptions based stratifications, we also recognize the possibility of expectation mismatches between users' self-conceptualization of their data and inferences on their data without consent or awareness [27]. For example if personalized ad allocations start working even better (by using momentary and contextual data), ads can seem "creepier" [55, 73] — as Malheiros et al. noted, "too personalized" ads can although catch more attention, but could also elicit discomfort about the personalization [60].

Further, an implication of our work is towards a future with fewer (but effective) ads — however, companies can misuse this opportunity as a business advantage to serve the same quantity of ads to generate more revenue — this calls for necessary ethical guidelines in place that limits maximum obtainable revenue per user as a function of their platform use. Taken together, researchers, ethicists, users, and platform designers together need to better establish the guidelines and standards of making data simultaneously useful and privacy-preserving. Future work into systems that move away from traditional profile-based targeting can support this ongoing discussion, in particular by offering an alternative that has so far been less explored and rarely used.

### 8.4 Limitations and Future Directions

Our work has limitations, some of which also suggest interesting future directions. We do not take content (e.g., what an ad is about) into account. While our work is a step towards understanding the role of context and momentary features, we note that future research can incorporate content to examine the interplay between context and content-centric factors in explaining online ad effectiveness. Additionally, our study functions within the limits of the existing ad allocation system on the platform. While our work assumed that platform-centric factors such as the content recommender algorithm's accuracy similarly applies to all users and any discrepancy is washed out in a large data, that may not be the case. Future work can control for the "goodness" of recommender algorithms if these metrics are available.

Our quasi-experimental approach only accounts for observed factors on the platform. Like any observational study, we cannot infer *true causality*. Future experimental studies based on our methodology and insights from our study can help confirm the validity and applicability of our findings. Similarly, we only quantify *observed* ad effectiveness, and cannot estimate the efficacy of the ads in terms of whether the users actually bought or used the products in the ads [66]. Future studies can enroll small representative samples of the population and conduct experimental studies that also incorporate the offline element of effectiveness of online ads. Future work can also study the "why"-s related to whether users like a particular kind of ad allocation versus the other [57]. Along the same lines, future studies can examine if providing explanation to ad allocations make users more (or less) conducive to ads [48]. We also cannot claim the generalizability of our findings on other platforms and other forms of ad allocations, which can be explored in future work. Our work builds the foundation for incorporating context and momentary behaviors in ad allocations, which can be extended in the future to other forms of content recommendation systems and problem settings.

## 9 CONCLUSION

This paper examined user ad consumption on online social platforms, particularly on the Snapchat Discover Feed. We conducted a quasi-experimental study on three months of longitudinal data of 100K Snapchat users. We split the longitudinal timeline of each user into Baseline and Measurement periods, where we operationalized "preferred timing" of ads based on ad reception in the Baseline period. Based on this, we obtained two groups of Treated and Control users based on how they were shown ads in the Measurement period. We conducted stratified propensity score analysis to match Treated and Control users by minimizing observed covariates such as aggregated activities and time spent on the platform. We found that timing ads at preferred times of users leads to effective ad outcomes (RTE>1.5). We then examined ad outcomes with respect to momentary activities on the platform, operationalized in terms of duration, interactivity, interaction diversity, extra-socialness, and distractedness. We made observations and recommendations related to ad allocations on preferred times and momentary on-platform behaviors. We simulated ad reallocation, finding that our study-driven insights lead to more valuable ad distributions. We also evaluated the robustness of our study design and parameter choices finding convergence in findings and validity to our study. We discussed the implications of our work in advancing our understanding of ad consumption on social media, and in designing better and responsible ad allocations from both user and platform perspectives.

## REFERENCES


[1] 2014. Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. *Journal of clinical epidemiology* 67, 3 (2014), 267–277.

[2] Paige Alfonzo. 2019. *Mastering mobile through Social Media: Creating engaging content on Instagram and Snapchat.* ALA TechSource.

[3] Muhammad Ali, Piotr Sapiezynski, Miranda Bogen, Aleksandra Korolova, Alan Mislove, and Aaron Rieke. 2019. Discrimination through Optimization: How Facebook's Ad Delivery Can Lead to Biased Outcomes. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–30.

[4] Maximilian Altmeyer, Kathrin Dernbecher, Vladislav Hnatovskiy, Marc Schubhan, Pascal Lessel, and Antonio Krüger. 2019. Gamified Ads: Bridging the Gap Between User Enjoyment and the Effectiveness of Online Ads. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.

[5] Aris Anagnostopoulos, Ravi Kumar, and Mohammad Mahdian. 2008. Influence and correlation in social networks. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 7–15.

[6] Nikola Banovic, Tofi Buzali, Fanny Chevalier, Jennifer Mankoff, and Anind K Dey. 2016. Modeling and understanding human routine behavior. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 248–260.

[7] Rajeev Batra and Douglas M Stayman. 1990. The role of mood in advertising effectiveness. *Journal of Consumer research* 17, 2 (1990), 203–214.

[8] Russell W Belk. 1974. An exploratory assessment of situational effects in buyer behavior. *Journal of marketing research* 11, 2 (1974), 156–163.

[9] Danah Boyd and Kate Crawford. 2012. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society* 15, 5 (2012), 662–679.

[10] Giorgio Brajnik and Silvia Gabrielli. 2010. A review of online advertising effects on the user experience. *International Journal of Human-Computer Interaction* 26, 10 (2010), 971–997.

[11] Jack W Brehm. 1966. A theory of psychological reactance. (1966).

[12] Laura Frances Bright and Kelty Logan. 2018. Is my fear of missing out (FOMO) causing fatigue? Advertising, social media fatigue, and the implications for consumers and brands. *Internet Research* (2018).

[13] Fred E Bronner, Jasper R Bronner, and John Faasse. 2007. In the mood for advertising. *International Journal of Advertising* 26, 3 (2007), 333–355.

[14] Jean-Louis Chandon, Mohamed Saber Chtourou, and David R Fortin. 2003. Effects of configuration and exposure levels in responses to web advertisements. *Journal of Advertising Research* 43, 2 (2003), 217–229.

[15] Jean-Charles Chebat, Francois Limoges, and Claire Gelinas-Chebat. 1997. Effects of circadian orientation, time of day, and arousal on consumers' depth of information processing of advertising. *Perceptual and motor skills* 85, 2 (1997), 479–490.

[16] Henriette Cramer. 2015. Effects of ad quality & content-relevance on perceived content quality. In *proceedings of the 33rd annual ACM conference on human factors in computing systems*. 2231–2234.

[17] Aron Culotta. 2014. Estimating county health statistics with Twitter. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1335–1344.

[18] Vedant Das Swain, Koustuv Saha, Hemang Rajvanshy, Anusha Sirigiri, Julie M Gregg, Suwen Lin, Gonzalo J Martinez, Stephen M Mattingly, Shayan Mirjafari, Raghu Mulukutla, et al. 2019. A Multisensor Person-Centered Approach to Understand the Role of Daily Activities in Job Performance with Organizational Personas. *Proc. IMWUT* (2019).

[19] Munmun De Choudhury, Emre Kiciman, Mark Dredze, Glen Coppersmith, and Mrinal Kumar. 2016. Discovering shifts to suicidal ideation from mental health content in social media. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2098–2110.

[20] Marco de Sa, Vidhya Navalpakkam, and Elizabeth F Churchill. 2013. Mobile advertising: evaluating the effects of animation, user and content relevance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2487–2496.

[21] Peter Doyle and John Saunders. 1990. Multiproduct advertising budgeting. *Marketing Science* 9, 2 (1990), 97–113.

[22] Robert H Ducoffe. 1995. How consumers assess the value of advertising. *Journal of Current Issues & Research in Advertising* 17, 1 (1995), 1–18.

[23] Robert H Ducoffe. 1996. Advertising value and advertising on the web-Blog@ management. *Journal of advertising research* 36, 5 (1996), 21–32.

[24] Robert H Ducoffe and Eleonora Curlo. 2000. Advertising value and advertising processing. *Journal of Marketing Communications* (2000).

[25] Julie A Edell and Marian C Burke. 1984. The moderating effect of attitude toward an ad on ad effectiveness under different processing conditions. *ACR North American Advances* (1984).

[26] Steven M Edwards, Hairong Li, and Joo-Hyun Lee. 2002. Forced exposure and psychological reactance: Antecedents and consequences of the perceived intrusiveness of pop-up ads. *Journal of advertising* 31, 3 (2002), 83–95.

[27] Casey Fiesler, Cliff Lampe, and Amy S Bruckman. 2016. Reality and perception of copyright terms of service for online content creation. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. 1450–1461.

[28] Meryl Paula Gardner. 1985. Mood states and consumer behavior: A critical review. *Journal of Consumer research* 12, 3 (1985), 281–300.

[29] Kiran Garimella, Orestis Kostakis, and Michael Mathioudakis. 2017. Ad-blocking: A study on performance, privacy and counter-measures. In *Proceedings of the 2017 ACM on Web Science Conference*. 259–262.

[30] Bob Gilbreath. 2017. Rise of Subscriptions and the Fall of Advertising: *https://medium.com/the-graph/rise-of-subscriptions-and-the-fall-of-advertising-d5e4d8800a49*.

[31] Scott A Golder and Michael W Macy. 2011. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science* 333, 6051 (2011), 1878–1881.

[32] Kendall Goodrich. 2013. Effects of age and time of day on Internet advertising outcomes. *Journal of Marketing Communications* (2013).

[33] Samuel D Gosling, Adam A Augustine, Simine Vazire, Nicholas Holtzman, and Sam Gaddis. 2011. Manifestations of personality in online social networks: Self-reported Facebook-related behaviors and observable profile information. *Cyberpsychology, Behavior, and Social Networking* 14, 9 (2011), 483–488.

[34] Kelley Gullo, Jonah Berger, Jordan Etkin, and Bryan Bollinger. 2019. Does time of day affect variety-seeking? *Journal of Consumer Research* 46, 1 (2019), 20–35.

[35] Qi Guo, Eugene Agichtein, Charles LA Clarke, and Azin Ashkan. 2009. In the mood to click? Towards inferring receptiveness to search advertising. In *2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*, Vol. 1. IEEE, 319–324.

[36] Hana Habib, Neil Shah, and Rajan Vaish. 2019. Impact of Contextual Factors on Snapchat Public Sharing. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.

[37] Edward L Hannan. 2008. Randomized clinical trials and observational studies: guidelines for assessing respective strengths and limitations. *JACC* (2008).

[38] Curt Haugtvedt, Richard E Petty, John T Cacioppo, and Theresa Steidley. 1988. Personality and ad effectiveness: Exploring the utility of need for cognition. *ACR North American Advances* (1988).

[39] Miguel A Hernan and James M Robins. 2010. *Causal inference.*

[40] Ben Hutchinson and Margaret Mitchell. 2019. 50 years of test (un) fairness: Lessons for machine learning. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*. 49–58.

[41] Guido W Imbens and Donald B Rubin. 2015. *Causal inference in statistics, social, and biomedical sciences.* Cambridge.

[42] Garrett A Johnson, Randall A Lewis, and Elmar I Nubbemeyer. 2017. Ghost ads: Improving the economics of measuring online ad effectiveness. *Journal of Marketing Research* 54, 6 (2017), 867–884.

[43] Jukka Jouhki, Epp Lauk, Maija Penttinen, Niina Sormanen, and Turo Uskali. 2016. Facebook's emotional contagion experiment as a challenge to research ethics. *Media and Communication* 4 (2016).

[44] Parisa Kaghazgaran, Maarten Bos, Leonardo Neves, and Neil Shah. 2020. Social Factors in Closed-Network Content Consumption. *CIKM* (2020).

[45] Komal Kapoor, Vikas Kumar, Loren Terveen, Joseph A Konstan, and Paul Schrater. 2015. "I like to explore sometimes" Adapting to Dynamic User Novelty Preferences. In *Proceedings of the 9th ACM Conference on Recommender Systems*. 19–26.

[46] Komal Kapoor, Karthik Subbian, Jaideep Srivastava, and Paul Schrater. 2015. Just in time recommendations: Modeling the dynamics of boredom in activity streams. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*. 233–242.

[47] Emre Kıcıman, Scott Counts, and Melissa Gasser. 2018. Using Longitudinal Social Media Analysis to Understand the Effects of Early College Alcohol Use.. In *ICWSM*. 171–180.

[48] Tami Kim, Kate Barasz, and Leslie K John. 2019. Why am I seeing this ad? The effect of ad transparency on ad effectiveness. *Journal of Consumer Research* 45, 5 (2019), 906–932.

[49] Yoojung Kim, Mihyun Kang, Sejung Marina Choi, and Yongjun Sung. 2016. To click or not to click? Investigating antecedents of advertisement clicking on Facebook. *Social Behavior and Personality: an international journal* 44, 4 (2016), 657–667.

[50] Gary King, Richard Nielsen, et al. 2016. Why propensity scores should not be used for matching. (2016).

[51] Farshad Kooti, Karthik Subbian, Winter Mason, Lada Adamic, and Kristina Lerman. 2017. Understanding short-term changes in online activity sessions. In *Proceedings of the 26th International Conference on World Wide Web Companion*. 555–563.

[52] Michal Kosinski, David Stillwell, and Thore Graepel. 2013. Private traits and attributes are predictable from digital records of human behavior. (2013).

[53] Adam DI Kramer, Jamie E Guillory, and Jeffrey T Hancock. 2014. Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences* 111, 24 (2014), 8788–8790.

[54] Hemank Lamba and Neil Shah. 2019. Modeling dwell time engagement on visual multimedia. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 1104–1113.

[55] Zhou Li, Kehuan Zhang, Yinglian Xie, Fang Yu, and XiaoFeng Wang. 2012. Knowing your enemy: understanding and detecting malicious web advertising. In *Proceedings of the 2012 ACM conference on Computer and communications security*. 674–686.

[56] Q Vera Liao, Wai-Tat Fu, and Sri Shilpa Mamidi. 2015. It is all about perspective: An exploration of mitigating selective exposure with aspect indicators. In *Proceedings of the 33rd annual ACM conference on Human factors in computing systems*. 1439–1448.

[57] Brian Y Lim, Anind K Dey, and Daniel Avrahami. 2009. Why and why not explanations improve the intelligibility of context-aware intelligent systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2119–2128.

[58] Jason Liu, Elissa R Weitzman, and Rumi Chunara. 2017. Assessing behavioral stages from social media data. In *CSCW*.

[59] Xun Lu and Halbert White. 2014. Robustness checks and robustness tests in applied economics. *Journal of econometrics* 178 (2014), 194–206.

[60] Miguel Malheiros, Charlene Jennett, Snehalee Patel, Sacha Brostoff, and Martina Angela Sasse. 2012. Too close for comfort: A study of the effectiveness and acceptability of rich-media personalized advertising. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 579–588.

[61] En Mao and Jing Zhang. 2015. What drives consumers to click on social media ads? The roles of content, media, and individual factors. In *2015 48th Hawaii International Conference on System Sciences*. IEEE, 3405–3413.

[62] En Mao and Jing Zhang. 2017. What affects users to click on display ads on social media? The roles of message values, involvement, and security. *Journal of Information Privacy and Security* 13, 2 (2017), 84–96.

[63] Gloria Mark, Shamsi T Iqbal, Mary Czerwinski, and Paul Johns. 2014. Bored mondays and focused afternoons: The rhythm of attention and online activity in the workplace. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 3025–3034.

[64] Jack Marshall. 2016. How to Persuade Consumers to Disable Ad Blockers: *https://www.wsj.com/articles/how-to-persuade-consumers-to-disable-ad-blockers-1469541611*.

[65] Akhil Mathur, Nicholas D Lane, and Fahim Kawsar. 2016. Engagement-aware computing: Modelling user engagement from mobile contexts. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 622–633.

[66] Daniel McDuff, Rana El Kaliouby, Jeffrey F Cohn, and Rosalind W Picard. 2014. Predicting ad liking and purchase intent: Large-scale analysis of facial responses to ads. *IEEE Transactions on Affective Computing* 6, 3 (2014), 223–235.

[67] Abhinav Mehrotra, Fani Tsapeli, Robert Hendley, and Mirco Musolesi. 2017. MyTraces: Investigating correlation and causation between users' emotional states and mobile phone interaction. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* (2017).

[68] Jacob Metcalf and Kate Crawford. 2016. Where are human subjects in big data research? The emerging ethics divide. *Big Data & Society* 3, 1 (2016), 2053951716650211.

[69] Jeremy Miles. 2014. Tolerance and variance inflation factor. *Wiley StatsRef: Statistics Reference Online* (2014).

[70] Ben Miroglio, David Zeber, Jofish Kaye, and Rebecca Weiss. 2018. The effect of ad blocking on user engagement with the web. In *Proceedings of the 2018 World Wide Web Conference*. 813–821.

[71] Bruce John Morrison and Marvin J Dainoff. 1972. Advertisement complexity and looking time. *Journal of marketing research* 9, 4 (1972), 396–400.

[72] Ngoc Thi Nguyen, Agustin Zuniga, Hyowon Lee, Pan Hui, Huber Flores, and Petteri Nurmi. 2020. (M) ad to See Me? Intelligent Advertisement Placement: Balancing User Annoyance and Advertising Effectiveness. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 2 (2020), 1–26.

[73] Katie O'Donnell and Henriette Cramer. 2015. People's perceptions of personalized ads. In *Proceedings of the 24th International Conference on World Wide Web*. 1293–1298.

[74] Alexandra Olteanu, Carlos Castillo, Fernando Diaz, and Emre Kiciman. 2019. Social data: Biases, methodological pitfalls, and ethical boundaries. *Frontiers in Big Data* 2 (2019), 13.

[75] Alexandra Olteanu, Onur Varol, and Emre Kiciman. 2017. Distilling the outcomes of personal experiences: A propensity-scored analysis of social media. In *Proc. CSCW*.

[76] Harshvardhan J Pandit and Dave Lewis. 2018. Ease and ethics of user profiling in black mirror. In *Companion Proceedings of the The Web Conference 2018*. 1577–1583.

[77] Pew. 2019. pewinternet.org/fact-sheet/social-media.

[78] Martin Pielot, Bruno Cardoso, Kleomenis Katevas, Joan Serrà, Aleksandar Matic, and Nuria Oliver. 2017. Beyond interruptibility: Predicting opportune moments to engage mobile phone users. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* (2017).

[79] Lin Qiu, Han Lin, Jonathan Ramsay, and Fang Yang. 2012. You are what you tweet: Personality expression and perception on Twitter. *Journal of research in personality* 46, 6 (2012), 710–718.

[80] Daniele Quercia, Michal Kosinski, David Stillwell, and Jon Crowcroft. [n.d.]. Our twitter profiles, our selves: Predicting personality with twitter.

[81] Manish Raghavan, Solon Barocas, Jon Kleinberg, and Karen Levy. 2020. Mitigating bias in algorithmic hiring: Evaluating claims and practices. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. 469–481.

[82] Vennila Ramalingam, B Palaniappan, N Panchanatham, and S Palanivel. 2006. Measuring advertisement effectiveness—a neural network approach. *Expert systems with applications* 31, 1 (2006), 159–163.

[83] Pei-Luen Patrick Rau, Qingzi Liao, and Cuiling Chen. 2013. Factors influencing mobile advertising avoidance. *International Journal of Mobile Communications* 11, 2 (2013), 123–139.

[84] Filipe N Ribeiro, Koustuv Saha, Mahmoudreza Babaei, Lucas Henrique, Johnnatan Messias, Fabricio Benevenuto, Oana Goga, Krishna P Gummadi, and Elissa M Redmiles. 2019. On microtargeting socially divisive ads: A case study of russia-linked ad campaigns on facebook. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*. 140–149.

[85] Steven Richmond. 2018. How Snapchat makes money. *Investopedia. Elérés* (2018).

[86] Helen Robinson, Anna Wysocka, and Chris Hand. 2007. Internet advertising effectiveness: the effect of design on click-through rates for banner ads. *International Journal of Advertising* 26, 4 (2007), 527–541.

[87] Christian Rohrer and John Boyd. 2004. The rise of intrusive online advertising and the response of user experience research at Yahoo!. In *CHI'04 Extended Abstracts on Human Factors in Computing Systems*. 1085–1086.

[88] Donald B Rubin. 2005. Causal inference using potential outcomes: Design, modeling, decisions. *J. Amer. Statist. Assoc.* 100, 469 (2005), 322–331.

[89] Adam Sadilek, Henry A Kautz, and Vincent Silenzio. 2012. Modeling Spread of Disease from Social Interactions.. In *International Conference on Weblogs and Social Media (ICWSM)*.
[90] Koustuv Saha and Amit Sharma. 2020. Causal Factors of Effective Psychosocial Outcomes in Online Mental Health Communities. In *ICWSM*.
[91] Koustuv Saha, Benjamin Sugar, John Torous, Bruno Abrahao, Emre Kıcıman, and Munmun De Choudhury. 2019. A Social Media Study on the Effects of Psychiatric Medication Use. In *ICWSM*.
[92] Koustuv Saha, Ingmar Weber, and Munmun De Choudhury. 2018. A Social Media Based Examination of the Effects of Counseling Recommendations After Student Deaths on College Campuses. In *ICWSM*.
[93] Ville Satopaa, Jeannie Albrecht, David Irwin, and Barath Raghavan. 2011. Finding a" kneedle" in a haystack: Detecting knee points in system behavior. In *ICDCS*.
[94] Ann E Schlosser, Sharon Shavitt, and Alaina Kanfer. 1999. Survey of Internet users' attitudes toward Internet advertising. *Journal of interactive marketing* 13, 3 (1999), 34–54.
[95] H Andrew Schwartz, Johannes C Eichstaedt, Margaret L Kern, et al. 2013. Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one* 8, 9 (2013), e73791.
[96] Yi Shen, Heshan Sun, Cheng Suang Heng, and Hock Chuan Chan. 2020. Facilitating Complex Product Choices on E-commerce Sites: An Unconscious Thought and Circadian Preference Perspective. *Decision Support Systems* (2020), 113365.
[97] Elizabeth A Stuart. 2010. Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics* 25, 1 (2010), 1.
[98] Gerard J Tellis, Rajesh K Chandy, Deborah MacInnis, and Pattana Thaivanich. 2005. Modeling the microeffects of television advertising: Which ad works, when, where, for how long, and why? *Marketing Science* (2005), 359–366.
[99] Catherine E Tucker. 2014. Social networks, personalized advertising, and privacy controls. *Journal of marketing research* 51, 5 (2014), 546–562.
[100] Daniel Tunkelang. 2018. Are Ads Really That Bad?: *https://medium.com/@dtunkelang/are-ads-really-that-bad-1c3d315f6689*.
[101] Richard Vaughn. 1980. How advertising works: A planning model. *Journal of advertising research* (1980).
[102] Aku Visuri, Simo Hosio, and Denzil Ferreira. 2017. Exploring mobile ad formats to increase brand recollection and enhance user experience. In *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia*. 311–319.
[103] Zack Whittaker. 2020. Apple's iOS 14 will give users the option to decline app ad tracking: *https://techcrunch.com/2020/06/22/apple-ios-14-ad-tracking*.
[104] Max Willer. 2018. New Data on Why People Hate Ads: Too Many, Too Intrusive, Too Creepy: *https://www.vieodesign.com/blog/new-data-why-people-hate-ads*.
[105] Max Willer. 2019. The Advertising Industry Has a Problem: People Hate Ads: *https://www.nytimes.com/2019/10/28/business/media/advertising-industry-research.html*.
[106] Stephan Winter, Ewa H Maslowska, and Anne L Vos. 2020. The effects of trait-based personalization in social media advertising. *Computers in Human Behavior* (2020), 106525.
[107] Lori D Wolin, Pradeep Korgaonkar, and Daulatram Lund. 2002. Beliefs, attitudes and behaviour towards Web advertising. *International Journal of Advertising* 21, 1 (2002), 87–113.
[108] Seounmi Youn and Seunghyun Kim. 2019. Understanding ad avoidance on Facebook: Antecedents and outcomes of psychological reactance. *Computers in Human Behavior* 98 (2019), 232–244.
[109] Zephoria. 2019. *https://zephoria.com/top-10-valuable-snapchat-statistics/*.
[110] Justine Zhang, Sendhil Mullainathan, and Cristian Danescu-Niculescu-Mizil. 2020. Quantifying the Causal Effects of Conversational Tendencies. *PACM HCI (CSCW)* (2020).